AD_____

Award Number: DAMD17-01-1-0221

TITLE: Characterization of IBC-1, a Novel Putative Breast Cancer Prognostic and Survival Factor

PRINCIPAL INVESTIGATOR: Minna Allinen, Ph.D.
Doctor Kornelia Polyak

CONTRACTING ORGANIZATION: Dana-Farber Cancer Institute
Boston, Massachusetts 02115

REPORT DATE: August 2004

TYPE OF REPORT: Annual Summary

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;
Distribution Unlimited

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.

20041118 096

BEST AVAILABLE COPY

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 074-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	August 2004	Annual Summary (1 Aug 01 - 31 Jul 04)

4. TITLE AND SUBTITLE
Characterization of IBC-1, a Novel Putative Breast Cancer Prognostic and Survival Factor

5. FUNDING NUMBERS
DAMD17-01-1-0221

6. AUTHOR(S)
Minna Allinen, Ph.D.
Doctor Kornelia Polyak

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Dana-Farber Cancer Institute
Boston, Massachusetts 02115

e-mail: kornelia_polyak@dfci.harvard.edu

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

12b. DISTRIBUTION CODE

13. Abstract ***(Maximum 200 Words)*** *(abstract should contain no proprietary or confidential information)*

Using SAGE (Serial Analysis of Gene Expression) we identified a SAGE tag that was only present in invasive breast carcinomas and their lymph node metastases. The transcript corresponding to this SAGE tag, IBC-1 (Invasive Breast Cancer-1)/dermcidin (DCD) encodes a secreted protein normally expressed only in the pons of the brain and sweat glands. Array CGH, FISH, and immunohistochemical analyses determined that IBC-1/DCD is overexpressed in approximately 10 % of invasive breast carcinomas, in some cases its overexpression is coupled with a focal copy number gain of its locus at 12q13.1, and its expression is associated with advanced clinical stage and poor prognosis. Expression of IBC-1/DCD in breast cancer cells promotes cell growth and survival, and reduces serum dependency. Putative high and low affinity receptors for IBC-1/DCD are present on the cell surface of breast carcinomas and neurons of the brain. Based on this data we hypothesize that IBC-1/DCD may play a role in tumorigenesis via enhancing cell growth and survival in a subset of breast carcinomas.

14. SUBJECT TERMS	15. NUMBER OF PAGES
None Provided	39
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

Table of Contents

Introduction

Breast cancer is one of the most common neoplasms in women and is a leading cause of cancer related deaths worldwide. Improved diagnostic tools have made it possible to detect breast cancers at early stages leading to a significant decrease in breast cancer mortality rates over the past decades. However, mortality rates of advanced stage cancer have not decreased significantly due to lack of effective therapies and approximately 25 % of breast cancer patients will die of their disease. Therefore, the development and application of new molecularly based diagnostic and prognostic tools and therapies is of utmost importance. The key to the development of such rational preventative and therapeutic approaches lies in the identification of genes and biochemical pathways involved in breast tumorigenesis. One approach to the discovery of novel diagnostic and prognostic markers, and therapeutic targets is to compare the gene expression profiles of normal and cancer cells and identify genes or subsets of genes with expression levels that correlate with tumor stage or clinical outcome. Several comprehensive gene expression profiling studies have been performed in breast cancer and several novel putative molecular markers have been identified. Most of these studies utilized array based platforms, and therefore, were inherently limited to the analysis of known genes and ESTs. SAGE is an alternative comprehensive gene expression profiling technique that does not require the a priori knowledge of the transcripts present in the cells. Thus, it allows for the identification of novel transcripts making it particularly suitable for the discovery of new molecular targets.

In this study we utilized the SAGE technology to determine the comprehensive gene expression profiles of normal breast tissue and breast carcinomas of all clinical stages with the aim of identifying genes involved in the initiation and progression of breast tumorigenesis. Based on this analysis we identified a novel gene, IBC-1 (Invasive Breast Carcinoma-1), encoding a small secreted protein that is only expressed in a subset of invasive breast carcinomas, in the pons of the brain, and in hippocampal neurons following oxidative insult, but not in any other normal human tissue. IBC-1 encodes a novel growth and survival factor that is overexpressed in a significant fraction of invasive breast carcinomas with poor prognostic features. IBC-1 has no significant homology to known proteins, therefore it may be involved in a novel-signaling pathway. Based on amino acid sequence IBC-1 corresponds to the cancer-associated cachexia-inducing factor previously identified by others.

- We developed highly sensitive ELISA assays using the different IBC-1 antibodies and determined that IBC-1 can be detected in the sera of breast cancer patients. We are currently collecting a larger set of samples to be tested in collaboration with Drs. Lyndsay Harris (DFCI) and (BU).

- We have generated a series of constitutive and inducible mammalian expression constructs, adenoviruses, and retroviruses expressing the human IBC-1 protein.

- We have determined that high and low affinity IBC-1 receptors are present on the cell surface of breast carcinomas and neurons of the brain. The presence of the high affinity receptor appears to correlate with cellular response to IBC-1.

- The expression or exogenous addition of IBC-1 to 21NT and SUM-52 breast cancer cell lines enhances cell proliferation and survival. We have also tested 5 additional breast cancer and immortalized cell lines, but IBC-1 had no effect on the growth of these other cells. Using ligand binding assays we determined that this could be due to the lack of a high-affinity IBC-1 receptor on these cells.

- We determined that IBC-1 has no effect on the 3D growth and invasion of 21NT and SUM52 cells (the cell lines that respond to IBC-1).

- We developed a TET-OFF inducible IBC-1 expression system in the MCFDCIS.com cell line and performed xenograft assays in nude mice. Based on one experiment we did not see an effect on tumor growth, but it appears that this cell line does not respond to IBC-1 in vitro neither. Unfortunately the 21NT and SUM52 cell lines are poorly tumorigenic, thus we have not been able to use them for xenograft experiments. We also tested if co-injection of human breast cancer stroma would enhance the tumorigenic potential of 21NT cells, but even using this approach we have not been able to generate xenografts.

- In collaboration with Dr. Giulio Passinetti (Mount Sinai, NY) we further characterized the possible involvement of IBC-1 in neurodegenerative disease. The results of these studies will be submitted for publication in the near future.

Reportable outcomes:

Please see attached manuscript published in Proc. Natl. Acad. Sci.. USA reporting the results of the IBC-1 study and an additional manuscript resulting from the work of Dr. Allinen currently in press in Cancer Cell.

Antibodies generated:

We have generated several different (2 N-terminal peptide and 1 C terminal peptide) rabbit polyclonal antibodies against the human IBC-1 protein and a variant of IBC-1 that is generated by alternative splicing and has a different C-terminus.

Cell lines generated:

We have generated several human cell lines that constitutively or inducibly overexpress the human IBC-1 protein or variant of IBC-1.

Conclusions:

In summary, we determined that IBC-1/DCD is a novel growth and survival factor that is overexpressed in approximately 10 % of primary invasive breast carcinomas and its overexpression at least in some cases is associated with a gain of its locus at 12q13.1. Based on its function and restricted expression pattern in normal adult tissues, IBC-1/DCD is a candidate cancer therapeutic target. The secreted nature and extracellular mechanism of DCD action makes it even more attractive for such a purpose.

Neurons are particularly sensitive to reactive oxygen species (ROS) whereas tumor cells themselves produce large amounts of ROS. Therefore, the high expression of IBC-1/DCD in these cell types may be essential for their survival. Thus, therapeutic activation of the IBC-1/DCD signaling pathway may be beneficial in certain neurodegenerative diseases involving catecholaminergic neurons such as Parkinson's disease, while its therapeutic inhibition may be an effective treatment of tumors with DCD expression.

References:

1. Greenlee, R. T., Murray, T., Bolden, S. & Wingo, P. A. (2000) *CA Cancer J Clin* **50,** 7-33.
2. Perou, C. M., Sorlie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., Rees, C. A., Pollack, J. R., Ross, D. T., Johnsen, H., Akslen, L. A., Fluge, O., Pergamenschikov, A., Williams, C., Zhu, S. X., Lonning, P. E., Borresen-Dale, A. L., Brown, P. O. & Botstein, D. (2000) *Nature* **406,** 747-52.
3. Sorlie, T., Perou, C. M., Tibshirani, R., Aas, T., Geisler, S., Johnsen, H., Hastie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., Thorsen, T., Quist, H., Matese, J. C., Brown, P. O., Botstein, D., Eystein Lonning, P. & Borresen-Dale, A. L. (2001) *Proc Natl Acad Sci U S A* **98,** 10869-74.
4. van de Vijver, M. J., He, Y. D., van't Veer, L. J., Dai, H., Hart, A. A., Voskuil, D. W., Schreiber, G. J., Peterse, J. L., Roberts, C., Marton, M. J., Parrish, M., Atsma, D., Witteveen, A., Glas, A., Delahaye, L., van der Velde, T., Bartelink, H., Rodenhuis, S., Rutgers, E. T., Friend, S. H. & Bernards, R. (2002) *N Engl J Med* **347,** 1999-2009.
5. van 't Veer, L. J., Dai, H., van de Vijver, M. J., He, Y. D., Hart, A. A., Mao, M., Peterse, H. L., van der Kooy, K., Marton, M. J., Witteveen, A. T., Schreiber, G. J., Kerkhoven, R. M., Roberts, C., Linsley, P. S., Bernards, R. & Friend, S. H. (2002) *Nature* **415,** 530-6.
6. Porter, D., Lahti-Domenici, J., Keshaviah, A., Bae, Y. K., Argani, P., Marks, J., Richardson, A., Cooper, A., Strausberg, R., Riggins, G. J., Schnitt, S., Gabrielson, E., Gelman, R. & Polyak, K. (2003) *Mol Cancer Res* **1,** 362-75.
7. Velculescu, V. E., Zhang, L., Vogelstein, B. & Kinzler, K. W. (1995) *Science* **270,** 484-7.
8. Polyak, K. & Riggins, G. J. (2001) *J Clin Oncol* **19,** 2948-58.
9. Kononen, J., Bubendorf, L., Kallioniemi, A., Barlund, M., Schraml, P., Leighton, S., Torhorst, J., Mihatsch, M. J., Sauter, G. & Kallioniemi, O. P. (1998) *Nat Med* **4,** 844-7.
10. Hulette, C. M., Welsh-Bohmer, K. A., Crain, B., Szymanski, M. H., Sinclaire, N. O. & Roses, A. D. (1997) *Arch Pathol Lab Med* **121,** 615-8.
11. Krop, I. E., Sgroi, D., Porter, D. A., Lunetta, K. L., LeVangie, R., Seth, P., Kaelin, C. M., Rhei, E., Bosenberg, M., Schnitt, S., Marks, J. R., Pagon, Z., Belina, D., Razumovic, J. & Polyak, K. (2001) *Proc Natl Acad Sci U S A* **98,** 9796-9801.
12. St Croix, B., Rago, C., Velculescu, V., Traverso, G., Romans, K. E., Montgomery, E., Lal, A., Riggins, G. J., Lengauer, C., Vogelstein, B. & Kinzler, K. W. (2000) *Science* **289,** 1197-202.
13. Qian, Y., Fritzsch, B., Shirasawa, S., Chen, C. L., Choi, Y. & Ma, Q. (2001) *Genes Dev* **15,** 2533-45.
14. Kuchinka, B. D., Kalousek, D. K., Lomax, B. L., Harrison, K. J. & Barrett, I. J. (1995) *Mod Pathol* **8,** 183-6.
15. Pinkel, D., Segraves, R., Sudar, D., Clark, S., Poole, I., Kowbel, D., Collins, C., Kuo, W. L., Chen, C., Zhai, Y., Dairkee, S. H., Ljung, B. M., Gray, J. W. & Albertson, D. G. (1998) *Nat Genet* **20,** 207-11.
16. Flanagan, J. G. & Leder, P. (1990) *Cell* **63,** 185-194.
17. Lal, A., Lash, A. E., Altschul, S. F., Velculescu, V., Zhang, L., McLendon, R. E., Marra, M. A., Prange, C., Morin, P. J., Polyak, K., Papadopoulos, N., Vogelstein, B., Kinzler, K. W., Strausberg, R. L. & Riggins, G. J. (1999) *Cancer Res* **59,** 5403-7.

18. Porter, D. A., Krop, I. E., Nasser, S., Sgroi, D., Kaelin, C. M., Marks, J. R., Riggins, G. & Polyak, K. (2001) *Cancer Res* **61,** 5697-702.
19. Burge, C. & Karlin, S. (1997) *J Mol Biol* **268,** 78-94.
20. Cariuk, P., Lorite, M. J., Todorov, P. T., Field, W. N., Wigmore, S. J. & Tisdale, M. J. (1997) *Br J Cancer* **76,** 606-13.
21. Cunningham, T. J., Hodge, L., Speicher, D., Reim, D., Tyler-Polsz, C., Levitt, P., Eagleson, K., Kennedy, S. & Wang, Y. (1998) *J Neurosci* **18,** 7047-60.
22. Schittek, B., Hipfel, R., Sauer, B., Bauer, J., Kalbacher, H., Stevanovic, S., Schirle, M., Schroeder, K., Blin, N., Meier, F., Rassner, G. & Garbe, C. (2001) *Nat Immunol* **2,** 1133-7.
23. Spitz, D. R., Sim, J. E., Ridnour, L. A., Galoforo, S. S. & Lee, Y. J. (2000) *Ann N Y Acad Sci* **899,** 349-62.
24. Sanghi, S., Kumar, R., Lumsden, A., Dickinson, D., Klepeis, V., Trinkaus-Randall, V., Frierson, H. F., Jr. & Laurie, G. W. (2001) *J Mol Biol* **310,** 127-39.
25. McDevitt, T. M. & Tisdale, M. J. (1992) *Br J Cancer* **66,** 815-20.
26. Todorov, P. T., McDevitt, T. M., Cariuk, P., Coles, B., Deacon, M. & Tisdale, M. J. (1996) *Cancer Res* **56,** 1256-61.
27. Todorov, P., Cariuk, P., McDevitt, T., Coles, B., Fearon, K. & Tisdale, M. (1996) *Nature* **379,** 739-42.
28. Cunningham, T. J., Jing, H., Akerblom, I., Morgan, R., Fisher, T. S. & Neveu, M. (2002) *Exp Neurol* **177,** 32-9.
29. Szatrowski, T. P. & Nathan, C. F. (1991) *Cancer Res* **51,** 794-8.

Appendices:

1. Porter D, Weremowicz S, Koei Chin, Seth P, Keshaviah A, Lahti-Domenici J, Bae YK, Monitto CL, Merlos-Suarez A, Chan J, Hulette CM, Richardson A, Morton CC, Marks J, Duyao M, Hruban R, Gabrielson E, Gelman R, and Polyak K. A neural growth factor is a candidate oncogene in breast cancer. Proc Natl Acad Sci. 2003; 100:10931-10936.

2. Allinen M, Beroukhim R, Cai, L, Brennan C, Lahti-Domenici J, Huang H, Porter D, Hu M, Chin L, Richardson A, Schnitt S, Sellers W, Polyak K. Molecular characterization of the tumor microenvironment in breast cancer. Cancer Cell 2004; *In press*

3. Seth P, Krop IE, Porter DA and Polyak K. Novel estrogen and tamoxifen induced genes identified by SAGE (Serial Analysis of Gene Expression). Oncogene 2002; 21:836-843.

A neural survival factor is a candidate oncogene in breast cancer

Dale Porter[a,b], Stanislawa Weremowicz[c,d], Koei Chin[e], Pankaj Seth[a,b], Aparna Keshaviah[f], Jaana Lahti-Domenici[a], Young Kyung Bae[g], Constance L. Monitto[h], Ana Merlos-Suarez[a,b], Jennifer Chan[c,d], Christine M. Hulette[i], Andrea Richardson[c,d], Cynthia C. Morton[d,j], Jeffrey Marks[i], Mabel Duyao[k], Ralph Hruban[g], Edward Gabrielson[g], Rebecca Gelman[f,l], and Kornelia Polyak[a,b,m]

Departments of [a]Medical Oncology and [f]Biostatistics, Dana–Farber Cancer Institute, Department of [c]Pathology and [j]Obstetrics, Gynecology, and Reproductive Biology, Brigham and Women's Hospital, and Departments of [b]Medicine and [d]Pathology, Harvard Medical School and [l]Harvard School of Public Health, Boston, MA 02115; [e]Comprehensive Cancer Center, University of California, San Francisco, CA 94143; Departments of [g]Pathology and [h]Anesthesiology and Critical Care Medicine, Johns Hopkins University School of Medicine, Baltimore, MD 21231; [i]Departments of Medicine and Pathology, Duke University Medical Center, Durham, NC 27710; and [k]Ardais Corporation, Lexington, MA 02421

Edited by Bert Vogelstein, The Sidney Kimmel Comprehensive Cancer Center at Johns Hopkins, Baltimore, MD, and approved July 23, 2003 (received for review May 17, 2003)

Using serial analysis of gene expression (SAGE), we identified a SAGE tag that was present only in invasive breast carcinomas and their lymph node metastases. The transcript corresponding to this SAGE tag, dermcidin (DCD), encodes a secreted protein normally expressed only in the pons of the brain and sweat glands. Array comparative genomic hybridization, fluorescence *in situ* hybridization, and immunohistochemical analyses determined that DCD is overexpressed in ≈10% of invasive breast carcinomas; in some cases its overexpression is coupled with a focal copy number gain of its locus at 12q13.1, and its expression is associated with advanced clinical stage and poor prognosis. Expression of DCD in breast cancer cells promotes cell growth and survival and reduces serum dependency. Putative high- and low-affinity receptors for DCD are present on the cell surface of breast carcinomas and neurons of the brain. Based on these data we hypothesize that DCD may play a role in tumorigenesis by means of enhancing cell growth and survival in a subset of breast carcinomas.

Breast cancer is one of the most common neoplasms in women and is a leading cause of cancer-related deaths worldwide. Improved diagnostic tools have made it possible to detect breast cancers at early stages, leading to a significant decrease in breast cancer mortality rates over the past decades (1). However, mortality rates of advanced-stage cancer have not decreased significantly because of a lack of effective therapies, and ≈25% of breast cancer patients will die of the disease (1). Therefore, the development and application of new molecularly based diagnostic and prognostic tools and therapies are of utmost importance. The key to the development of such rational preventive and therapeutic approaches lies in the identification of genes and biochemical pathways involved in breast tumorigenesis. One approach to the discovery of novel diagnostic and prognostic markers and therapeutic targets is to compare the gene expression profiles of normal and cancer cells and identify genes or subsets of genes with expression levels that correlate with tumor stage or clinical outcome. Several comprehensive gene expression profiling studies have been performed in breast cancer, and several novel putative molecular markers have been identified (2–6). Most of these studies used array-based platforms and, therefore, were inherently limited to the analysis of known genes and ESTs. Serial analysis of gene expression (SAGE) is an alternative comprehensive gene expression profiling technique that does not require the *a priori* knowledge of the transcripts present in the cells. Thus, it allows for the identification of novel transcripts, making it particularly suitable for the discovery of new molecular targets (7, 8).

In this study we used the SAGE technology to determine the comprehensive gene expression profiles of normal breast tissue and breast carcinomas of all clinical stages with the aim of identifying genes involved in the initiation and progression of breast tumorigenesis. This approach led to the identification of a previously uncharacterized growth and survival factor that is overexpressed in a significant fraction of invasive breast carcinomas with poor prognostic features.

Methods

Cell Lines and Tissue Specimens. Breast cancer cell lines were obtained from American Type Culture Collection or were generously provided by Steve Ethier (University of Michigan, Ann Arbor), Gail Tomlinson (University of Texas, Austin), and Arthur Pardee (Dana–Farber Cancer Institute). Cells were grown in media recommended by the provider. Tumor specimens were obtained from Brigham and Women's and Massachusetts General Hospitals (Boston), Duke University, University Hospital Zagreb (Zagreb, Croatia), and the National Disease Research Interchange, snap frozen on dry ice, and stored at −80°C until use. All human tissue was collected using protocols approved by the institutional review boards. Tissue microarrays were (*i*) obtained from Imgenex (San Diego), Ambion (Austin, TX), Ardais Corporation, and Gentaur (Brussels); (*ii*) provided by the Cooperative Breast Cancer Tissue Resource; and (*iii*) generated at Johns Hopkins University and at Beth Israel Deaconess Medical Center following published protocols (9). Brain samples were collected from autopsies performed at Brigham and Women's Hospital and from subjects prospectively enrolled in the Rapid Autopsy Program of the Joseph and Kathleen Price Bryan Alzheimer Disease Research Center at Duke University Medical Center (10).

RNA Preparation, mRNA *in Situ* Hybridization, and Northern Blot Analysis. We performed RNA isolation, RT-PCR, and Northern blot analyses as described (11). We performed mRNA *in situ* hybridization using paraffin sections and digitonin-labeled riboprobes following a protocol developed by St. Croix *et al.* (12), and we hybridized frozen sections as described with minor modifications (13).

Dermcidin (DCD) Expression in Mammalian Cells and Growth and Survival Assays. We generated an N-terminal alkaline phosphatase (AP) C-terminal DCD fusion protein using the AP-TAG-5

This paper was submitted directly (Track II) to the PNAS office.

Abbreviations: AP, alkaline phosphatase; FISH, fluorescence *in situ* hybridization; BAC, bacterial artificial chromosome; CGH, comparative genomic hybridization; DCD, dermcidin; IBC-1, invasive breast cancer 1; ROS, reactive oxygen species; SAGE, serial analysis of gene expression.

[m]To whom correspondence should be addressed. E-mail: kornelia_polyak@dfci.harvard.edu.

© 2003 by The National Academy of Sciences of the USA

expression vector (GenHunter, Nashville, TN). We transfected mammalian cells with FuGENE6 (Roche), Lipofectamine, or Lipofectamine 2000 (Life Technologies, Rockville, MD) reagents. For mammalian expression, we subcloned the human DCD cDNA into the pBabe construct and confirmed DCD protein expression by immunoblot analysis. To determine the effect of DCD expression on cell growth, we plated 5,000 control (pBabe) or DCD-expressing (pBabe-DCD) cells per well in a 24-well plate, and 21NT cells were grown in either complete MCF10A medium (American Type Culture Collection) or MCF10A medium diluted 1:10 with basal MCF10A medium without growth factors added. Cells were counted (three wells per time point) on days 1, 3, 5, and 7 after plating. For menadione survival assays, 21NT pBabe and 21NT pBabe-DCD stable pools were plated (10^5 cells per well in a 24-well plate). At 6 h, cells were washed and medium was changed to serum-free DMEM-F12 medium with or without menadione (0, 100, and 200 μm; three wells per treatment), and cells were counted at 24 h. The experiment was repeated three times. For glucose deprivation assays, 21NT pBabe and 21NT pBabe-DCD stable pools were plated (5×10^4 cells per well in a 24-well plate). At 6 h, cells were washed and medium was changed to basal media with 5% or 0.5% horse serum and 0 or 4 mM glucose (three wells per treatment), and cells were counted at 48 h. The experiment was repeated three times.

Fluorescence *in Situ* Hybridization (FISH) and Array Comparative Genomic Hybridization (CGH). We obtained bacterial artificial chromosomes (BACs) containing the human *DCD*, *CDK4*, *SAS*, *GLI*, and *MDM2* genes from Research Genetics (Huntsville, AL). The d1223 probe for identification of chromosome 12 was obtained from Vysis (Naperville, IL). We performed FISH to paraffin-embedded or frozen tissue and metaphase chromosomes from normal human lymphocytes as described (14). We performed BAC array CGH essentially as described (15). DNA copy number variations that deviated significantly (at least three times higher than the standard deviation of the overall fluorescence intensity of the tumor DNA) from background ratios measured in normal genomic DNA control hybridizations were considered real copy number variations. In the case of the BAC containing DCD, the average log fluorescence ratio was 0.3. The detailed results of the array CGH analysis of the 152 breast tumors will be reported elsewhere.

Antibodies and Immunoblot, Immunohistochemical, and Statistical Analyses. We generated an affinity-purified polyclonal anti-DCD antibody against a synthetic peptide (RQAPKPRKQRSS) corresponding to amino acids 53–64 of the human DCD protein (Zymed), and other antibodies used were obtained from sources previously described (6). We performed immunoblot analyses as described (11). We analyzed the expression of DCD in primary tumors by the use of immunohistochemistry to tissue microarrays that contained evaluatable paraffin-embedded specimens derived from ductal carcinoma *in situ*, primary invasive breast cancer and distant breast cancer metastases, pancreatic, gastric, prostate, kidney, and colon carcinomas, melanomas, lymphomas, and gliomas. Immunohistochemical and statistical analyses were performed as described (6).

Ligand Binding Assays. We performed *in vivo* and *in vitro* ligand binding assays on primary tissues and cell lines using AP-DCD essentially as described (16). Briefly, we fixed frozen sections of various human specimens, incubated with either AP-DCD fusion protein or AP control-conditioned medium, rinsed, and then incubated with AP substrate forming a blue/purple precipitate. For *in vitro* assays we incubated cells in suspension with conditioned medium containing either AP alone or AP-DCD fusion protein, rinsed, and then assayed for bound AP activity.

Results

Identification of Invasive Breast Cancer 1 (IBC-1)/DCD. To identify genes implicated in breast tumorigenesis we determined the gene expression profiles of normal mammary epithelial cells and *in situ*, invasive, and metastatic breast carcinomas using SAGE. Using this approach we identified a SAGE tag with no database match that was highly expressed only in a subset of invasive breast carcinomas (17, 18) designated IBC-1. Searching the human genome sequence with the IBC-1 SAGE tag and 5′ *Nla*III site (5′-CATGACGTTAAAGAC-3′), we identified a genomic clone containing this tag and predicted (19) that it encodes a transcribed gene composed of five exons (Fig. 1*A*). Confirming the restricted expression pattern suggested by SAGE, based on Northern blot hybridization IBC-1 was expressed in only two regions of the brain: in the pons and, at a lower level, in the paracentral gyrus of the cerebral cortex, and not in 75 other normal human adult and fetal tissues (Fig. 1*B*). The predicted IBC-1 gene encodes a 110-aa protein with limited homology to lacritin and an EST containing an N-terminal signal peptide (Fig. 1 *C* and *D*). Further database searches using the predicted IBC-1 protein sequence revealed that IBC-1 nearly matches a 20-aa peptide derived from the mouse proteolysis-inducing factor or cachectic factor, and exactly matches a 30-aa neural survival-promoting peptide (20, 21) (Fig. 1*B*). While this work was in progress another group independently identified a cDNA from human sweat glands identical to IBC-1 and named it DCD (22). Thus, to avoid confusion due to multiple gene names, we renamed IBC-1 as DCD.

Expression of DCD in Breast Carcinomas and Correlation with Histopathologic Features. Next, we analyzed the expression of DCD in normal breast organoids, primary breast carcinomas, and breast cancer cell lines by Northern blot, RT-PCR, and mRNA *in situ* hybridization analyses and determined that it was expressed only in a subset of breast cancer cell lines and primary tumors (Fig. 2 *A–C* and data not shown). To determine the expression of DCD at the cellular level we performed mRNA *in situ* hybridization. Intense red or black (depending on hybridization protocol used) staining demonstrates that DCD is expressed in tumor cells and not in stromal cells (Fig. 2*C*). No signal was observed in adjacent normal mammary epithelial cells (Fig. 2*C*). In tumors 15 and 238 only a subset of cells showed high DCD expression indicating intratumoral heterogeneity (Fig. 2*C*).

To evaluate the expression of the DCD protein we performed immunohistochemical analysis of several tissue microarrays composed of breast carcinomas (Fig. 2*D*). Correlating with our SAGE results we detected DCD expression in primary invasive breast carcinomas (48/558), and rarely in ductal carcinoma *in situ* (1/70) or distant metastases (1/49). Statistical analysis determined that DCD-positive breast tumors were more likely to be of advanced stage (tumor node metastasis stage 2 or 3, mostly due to higher T and N, $P = 0.007$) indicating that DCD expression correlates with larger tumor size and with the presence of metastatic lymph nodes. Because both of these tumor characteristics are known to predict a bad prognosis, we analyzed DCD expression in relation to overall and distant metastasis-free survival in a subset of breast tumors with clinical follow-up data. Patients with DCD-positive tumors appeared to have decreased overall and distant metastasis-free survival, but this decrease did not reach statistical significance (data not shown).

We also analyzed DCD expression in multiple human tumor types and found that 2/64 pancreatic carcinomas expressed DCD. Thus, DCD overexpression may occur in other human tumor types, but the determination of this will require the examination of large tumor sets from each tumor type. Although the staining of melanomas did not detect any DCD-positive tumor cells, adjacent sweat glands of the skin were strongly

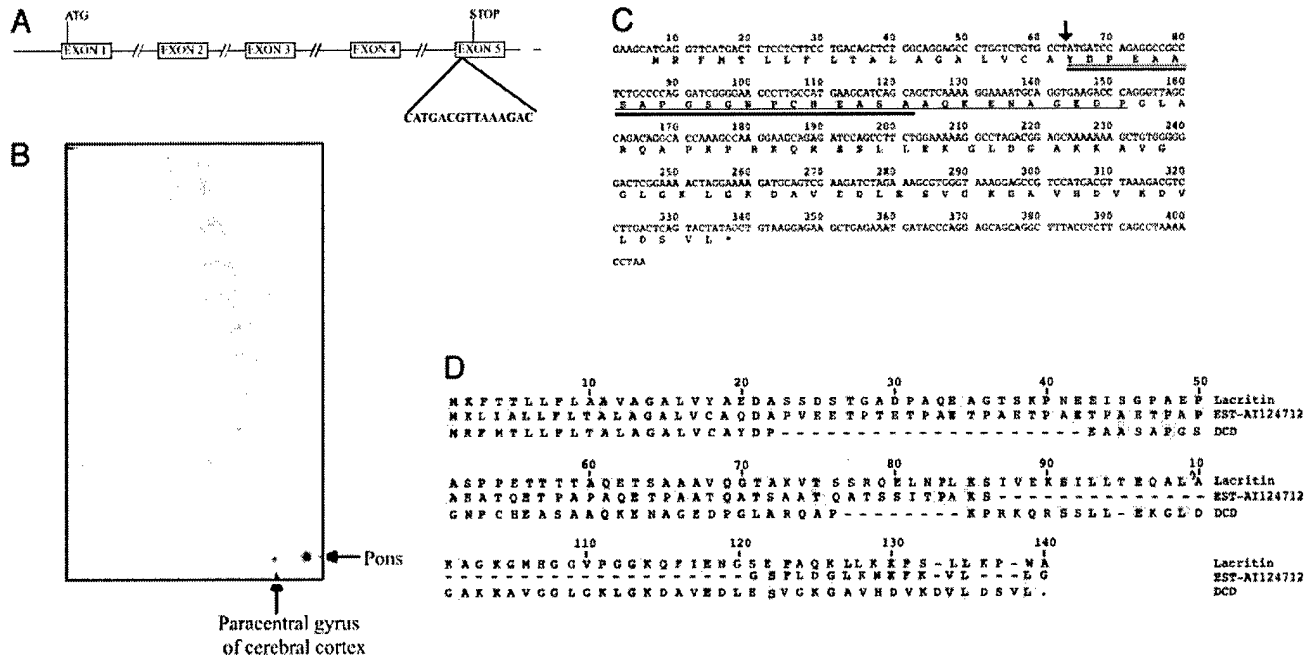


Fig. 1. DCD and its homologues. (*A*) Genomic structure of the human IBC-1/DCD gene. Exon–intron boundaries, start-and-stop codons, and the SAGE tag that led to the identification of IBC-1/DCD are indicated. (*B*) Evaluation of DCD expression in 76 human adult and fetal tissues on a dot-blot expression array. High level of expression was detected only in the pons of the brain, whereas low-level expression was seen in the paracentral gyrus of cerebral cortex. (*C*) Human IBC-1/DCD cDNA and predicted amino acid sequence. Sequences of the peptides derived from cachectic factor and survival peptide are indicated by thick and thin underlining, respectively. An arrow marks the predicted secretory signal peptidase cleavage site. (*D*) Amino acid alignment of DCD, lacritin, and EST-AI12471 proteins. Amino acids identical to the consensus are shaded in gray. Comparison was made by using DNAStar and the CLUSTAL algorithm.

DCD-positive (Fig. 2*D*), confirming DCD expression in sweat glands (22).

Despite an extensive analysis of cell lines from various tumor types, we were not able to identify a cell line that endogenously expresses the DCD protein at levels detectable by Western blot analysis using our antibody (data not shown). Thus, to confirm that the DCD transcript we identified encodes a protein that exists *in vivo*, we performed immunoblot analysis of DCD-

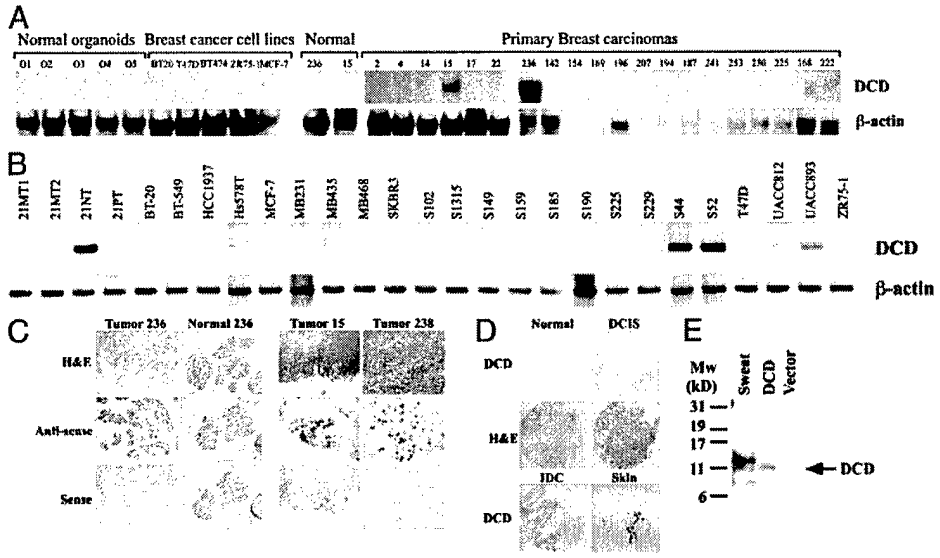
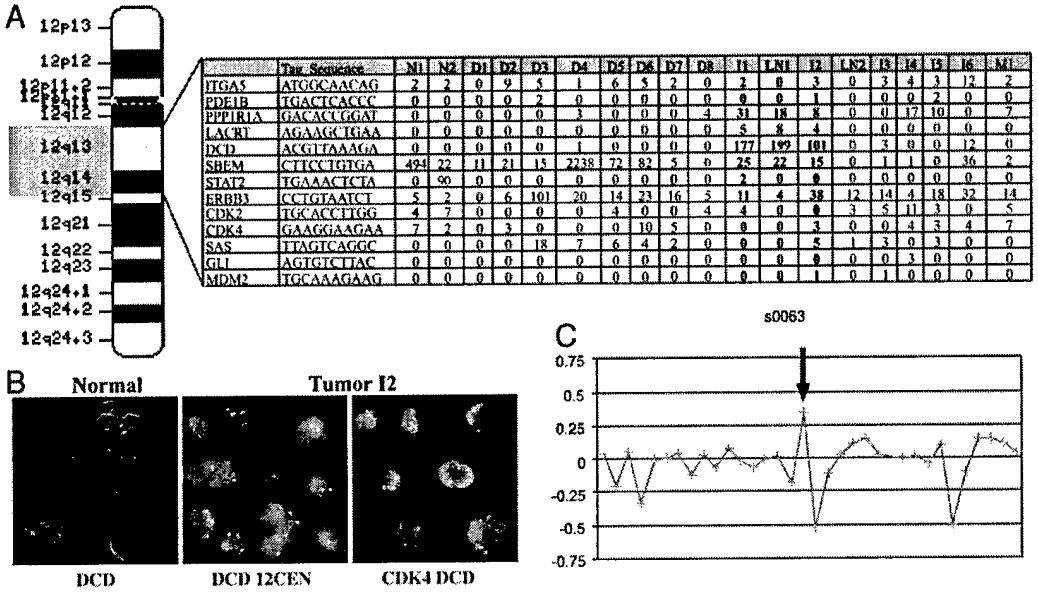


Fig. 2. Expression of DCD in normal and cancerous tissues. (*A*) Northern blot analyses of normal breast organoids, breast cancer cell lines, primary breast carcinomas, and corresponding normal breast tissue. High DCD expression is detected in only two tumors (15 and 236). The blot was rehybridized with β-actin to indicate equal loading. (*B*) RT-PCR analyses of breast cancer cell lines using DCD- and β-actin-specific primers. (*C*) mRNA *in situ* hybridization using digitonin-labeled DCD riboprobes on tissue sections (tumor and normal 236, red staining; tumors 15 and 238, black staining). Adjacent section stained with hematoxylin/eosin. (*D*) DCD immunostaining of normal breast tissue, ductal carcinoma *in situ*, a DCD-positive invasive breast carcinoma (IDC), and sweat gland of the skin. (*E*) Immunoblot analysis of human sweat and cells transfected with empty or DCD-expressing vector. An ≈11-kDa protein is detected in both transfected cells and sweat.

MEDICAL SCIENCES



	Tag Sequence	N1	N2	D1	D2	D3	D4	D5	D6	D7	D8	I1	LN1	I2	LN2	I3	I4	I5	I6	M1
ITGA5	ATGGCAACAG	2	2	0	9	5	1	6	5	2	0	2	0	3	0	3	4	3	12	2
PDE1B	TGACTCACCC	0	0	0	0	2	0	0	0	0	0	0	0	1	0	0	0	2	0	0
PPP1R1A	GACACCGGAT	0	0	0	0	0	3	0	0	0	4	31	18	8	0	0	17	10	0	7
LACRT	AGAAGCTGAA	0	0	0	0	0	0	0	0	0	0	5	8	4	0	0	0	0	0	0
DCD	ACGTTAAAGA	0	0	0	0	0	1	0	0	0	0	177	199	101	0	3	0	0	12	0
SBEM	CTTCCTGTGA	494	22	11	21	15	2238	72	82	5	0	25	22	15	0	1	1	0	36	2
STAT2	TGAAACTCTA	0	90	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0
ERBB3	CCTGTAATCT	5	2	0	6	101	20	14	23	16	5	11	4	38	12	14	4	18	32	14
CDK2	TGCACCTTGG	4	7	0	0	0	0	4	0	0	4	4	0	0	3	5	11	3	0	5
CDK4	GAAGGAAGAA	7	2	0	3	0	0	0	10	5	0	0	0	3	0	0	4	3	4	7
SAS	TTAGTCAGGC	0	0	0	0	18	7	6	4	2	0	0	0	5	1	3	0	3	0	0
GLI	AGTGTCTTAC	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0
MDM2	TGCAAAGAAG	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0

Fig. 3. Gain of DCD locus in breast carcinomas. (*A*) Idiogram of human chromosome 12 and the expression of genes adjacent to DCD in SAGE libraries generated from normal (N1and N2) breast tissue, and *in situ* (D1–8), invasive (I1–6), and metastatic (LN1–2 and M1) breast carcinomas. Genes closest to DCD (highlighted with yellow color), lacritin (LACRT), and, to a lesser extent, a phosphatase subunit (PPP1R1A) are expressed only in the three tumors with high levels of DCD, suggesting possible amplification of this chromosomal area. No other genes near DCD appear to be overexpressed in these breast tumors. (*B*) FISH analysis of DCD to normal metaphase chromosomes shows hybridization at 12q13 on both copies of chromosome 12 (*Left*). Hybridization of DCD (red) and an alpha-satellite probe to the centromere of chromosome 12 (green) reveals amplification of DCD and disomy of chromosome 12 in tumor I2 interphase cells (*Center*). Analysis of DCD (green) and CDK4 (red) reveals coamplification in tumor I2 interphase cells (*Right*). (*C*) A representative BAC array CGH profile demonstrating a gain of the DCD locus (arrow).

transfected cells and human sweat. Correlating with its predicted molecular weight, the exogenously expressed recombinant DCD protein migrates as a single ≈11-kDa protein, and a protein of approximately the same size is also detected in sweat (Fig. 2*E*). The slightly higher and lower molecular weight proteins recognized with our DCD antibody in the sweat may correspond to posttranslationally modified or partially proteolyzed DCD (Fig. 2*E*). These results confirm that a full-length DCD protein is expressed and secreted *in vivo*.

Focal Copy Number Gain of the DCD Locus in Breast Carcinomas. Based on the human genome sequence, DCD was localized to chromosome 12 in band q13.1, which we confirmed by FISH (Fig. 3*B*). Examination of the expression of all known and predicted genes in the vicinity (±5 megabases) of DCD determined that two genes localized next to DCD were expressed only in the same three breast carcinomas that expressed DCD and were not detected in any of the other >100 SAGE libraries (Fig. 3*A* and data not shown). This suggested that the overexpression of DCD in breast tumors may be due to gene amplification. To determine whether the DCD locus is amplified in the DCD-overexpressing tumors we performed FISH and detected moderate levels of DCD amplification in tumor I2 (Fig. 3*B*). We also analyzed several known oncogenes (*CDK4*, *SAS*, *GLI1*, and *MDM2*) localized to 12q and detected only CDK4 amplification in tumor I2 (Fig. 3*B* and data not shown). In three other tumors (I1, LN1, and 236) overexpressing DCD the FISH pattern was consistent with three to five copies of DCD and all of the other genomic regions tested, suggesting that a large part of chromosome 12 was gained (data not shown). However, based on SAGE, these oncogenes (*MDM2*, *CDK4*, *SAS*, etc.) were not overexpressed in DCD-positive breast tumors (Fig. 3*A*). To establish how frequently a gain of the DCD locus is detected in breast tumors, we analyzed an independent set of 152 breast tumors by using BAC array CGH and found a significant focal copy number increase of the DCD locus in 20 tumors (Fig. 3*C*).

To further investigate the association between gain of the DCD locus and the overexpression of the DCD protein, we performed immunohistochemical analysis on eight breast tumors that showed a 12q13.1 focal copy number increase of the DCD locus based on BAC array CGH. Five of these eight tumors expressed the DCD protein, which is a much higher fraction than expected ($P = 0.0003$) based on the frequency of DCD positivity in unselected tumors (48/558). Thus, this result further suggests that at least in some cases DCD overexpression in breast tumors is due to a gain of the 12q13.1 chromosomal area.

DCD Is a Growth and Survival Factor for Breast Cancer Cells. To analyze the effect of DCD overexpression on breast cancer cell growth and survival we established derivatives of the 21NT breast cancer cell line, chosen based on its features resembling DCD-expressing primary breast tumors, that stably overexpressed DCD. Next we compared the growth of pools of control, empty vector transfected cells with that of cells expressing DCD and found that DCD-expressing 21NT cells grew significantly faster than controls, especially in reduced serum-containing medium (Fig. 4*A*). Similar results were obtained in DCD-expressing VA13-transformed fibroblasts and C2C12 myoblasts, whereas preliminary data suggest that DCD has no effect on immortalized mammary epithelial cells (data not shown).

To determine the effect of DCD expression on cell survival after oxidative stress, we treated control and DCD-expressing 21NT cells with varying concentrations of menadione, a potent inducer of mitochondrial reactive oxygen species (ROS) production. As depicted in Fig. 4*B Left*, DCD-expressing cells were significantly more resistant to menadione-induced cell death than control 21NT cells. To establish whether the DCD-mediated protection from ROS-induced cell death is important in a more physiologic oxidative stress-inducing condition, we

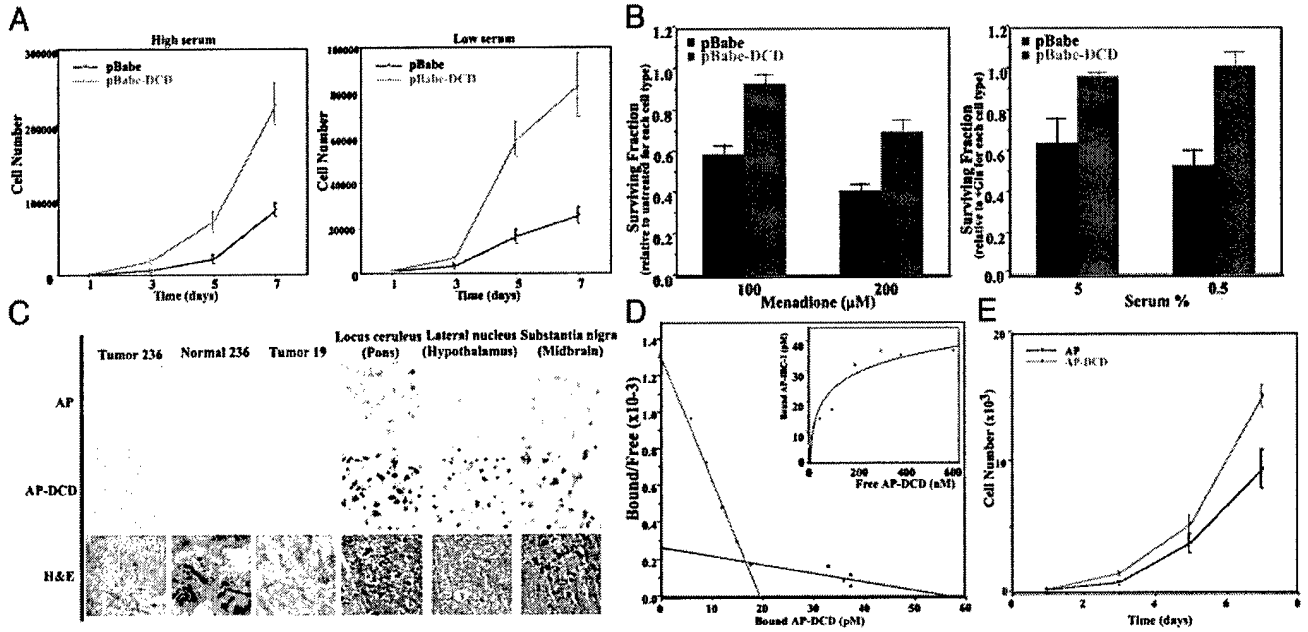


Fig. 4. DCD function and receptors. (*A*) Growth curves of control (pBabe) and DCD-expressing (pBabe-DCD) 21NT breast cancer cells in high (5%) or low (0.5%) serum-containing media. DCD-expressing cells grew significantly faster in both conditions. A representative experiment is shown. (*B*) Survival data showing reduced susceptibility of 21NT cells expressing DCD to cell death induced by menadione or glucose deprivation. Data are the mean of three experiments with three determinations each. (*C*) *In situ* staining for DCD receptor in breast and brain tissue. Sections of breast tumors, normal breast tissue, and brain were incubated with AP control or AP-DCD fusion protein. Purple staining detects the presence of a putative DCD receptor. Faint brownish coloring of neurons of the locus ceruleus and substantia nigra in the control AP sections is due to natural pigment (melanin) present in these cells. (*D*) Scatchard transformation of binding analysis of AP-DCD to 21NT breast cancer cells. (*Inset*) The actual binding curve. (*E*) Growth curves of cells treated with purified AP-DCD.

analyzed the effect of glucose deprivation on control 21NT and DCD-expressing cells. Cancer cells are known to be particularly sensitive to the withdrawal of glucose that leads to increased mitochondrial ROS production and subsequent cell death (23). Similar to the results obtained with menadione, DCD-expressing 21NT cells survived growth in glucose-free medium significantly better than control cells, with the most pronounced difference seen in low-serum-containing medium (Fig. 4*B Right*).

Cell Surface DCD Binding. The DCD protein is predicted to be secreted, suggesting that DCD is likely to execute its function through binding to a cell surface receptor. To determine whether there is a DCD-binding cell surface protein(s), we generated an AP-DCD fusion protein to be used as a ligand in receptor binding assays (16). Conditioned medium containing AP-DCD or control AP was used to stain normal and cancerous mammary tissue sections. Intense purple staining indicated the presence of a DCD-binding protein in tumor 236, but not in normal mammary epithelial and stromal cells, whereas low-intensity staining was observed in tumor 19 (Fig. 4*C*). These results suggested the presence of a cell surface DCD-binding protein(s) in cancerous, but not normal, mammary epithelial cells, and are consistent with an autocrine and/or paracrine mechanism of DCD action.

Because of its expression pattern in normal human brain, we also tested whether neurons bind DCD (Fig. 4*C*). Weak DCD binding to almost all neurons was seen in human adult brain (data not shown), whereas the strongest DCD binding was detected in neurons of the locus ceruleus, nucleus raphe pontis, substantia nigra, and the lateral hypothalamic nuclei (Fig. 4*C*).

To further test the binding characteristics of AP-DCD, we performed *in vitro* ligand binding assays using various cell lines. Low-level AP-DCD binding was detected in all cell lines tested, with stronger binding observed in human 21NT breast cancer cells (data not shown). To further characterize the AP-DCD-putative DCD-receptor interaction, we performed more detailed binding assays in 21NT breast cancer cells. Scatchard plot analysis showed two binding slopes in 21NT cells (Fig. 4*D*): one with a moderately high affinity ($K_d = 1.5 \times 10^{-8}$ M) and another with much lower affinity ($K_d = 2.1 \times 10^{-7}$ M). Further proving that DCD's effect is mediated through a cell surface receptor and that the AP-DCD fusion protein is a functional ligand for the putative DCD receptor, 21NT cells incubated in conditioned medium containing AP-DCD, or treated with purified AP-DCD, grew faster than controls (Fig. 4*E* and data not shown).

Discussion

Based on SAGE analysis of breast tumors of different clinical stages we identified DCD, a novel growth and survival factor for breast cancer cells. DCD encodes a secreted protein with limited homology to lacritin and an EST. Lacritin is a secretion-enhancing and growth-promoting factor recently identified from lacrimal gland (24). The EST is expressed in the cerebral cortex, and it encodes an uncharacterized protein containing a repetitive sequence (ETPA) found in several secreted proteins. In addition, two small proteolytic peptides identified as a cancer cachexia factor and a neural survival-promoting peptide, respectively, were likely to be derived from DCD (20, 21). The cachexia- and proteolysis-inducing factor was identified as a 24-kDa glycoprotein produced by the cachexia-inducing MAC 16 murine colon adenocarcinoma in mice and was shown to be present in the urine of cachectic cancer patients (25–27). The neural survival-promoting peptide was identified from the media of mouse HN33.1 hippocampal neurons and human Y79 retinoblasts treated with hydrogen peroxide and was shown to enhance neural survival after an oxidative insult (21, 28).

Based on FISH and array CGH analysis we determined that the overexpression of DCD in breast tumors is due to the focal copy number increase of the DCD locus. The low level of gain

observed in array CGH could be due to the fact that the breast tumors used for this analysis were not microdissected; thus, contaminating stromal cells with two copies of the DCD locus may decrease the hybridization signal. In addition, as depicted in Figs. 2*C* and 3*B*, the expression and gain of DCD are heterogeneous in most breast tumors, with only a fraction of tumor cells being positive; thus, even a significant copy number gain may be detected only as a low level gain when the tissue is analyzed in bulk by using array CGH. Correlating with this, cDNA array CGH analysis of the tumors with significant DCD gain based on FISH revealed no significant copy number increase (data not shown). The minimum region of chromosomal gain based on BAC aCGH is 4.2 megabases, because the two flanking BACs that do not show gain are that far apart. This region encompasses *DCD*, *CDK4*, and *SAS*. However, based on SAGE, we did not see overexpression of any of these oncogenes (*MDM2*, *CDK4*, *SAS*, etc.) in DCD-positive breast tumors (Fig. 3*A*).

Consistent with being a putative oncogene, the overexpression of DCD in breast cancer cells enhanced cell growth and survival and reduced serum dependency. Because in the cell survival experiments performed with menadione cell viability was determined 24 h after plating the cells, the observed difference seen in live cell numbers is unlikely to be the effect of DCD on cell growth. Conversely, the effect of DCD on cell growth cannot fully be explained by its ability to protect against ROS generated because of culturing the cells under supraphysiologic (21% atmospheric) oxygen concentrations, as a similar effect was seen in cells grown at physiologic (3%) oxygen concentration (data not shown). Thus, the growth and survival-promoting effects of DCD appear to be distinct, although to conclusively prove this will require the detailed characterization of the DCD-signaling pathway. Correlating with the observed *in vitro* effect of DCD on breast cancer cell growth and survival, DCD-expressing primary breast tumors were larger and more likely to have metastatic lymph nodes. Based on these results, it is likely that the overexpression and copy number gain of DCD confer a selective advantage for breast tumor cells.

Based on *in vivo* ligand binding studies performed with an AP-DCD fusion protein, we detected strong cell surface DCD binding to breast cancer cells and neurons of the brain. Interestingly, catecholaminergic (noradrenergic and dopaminergic) neurons of the brain that strongly bound AP-DCD are particularly susceptible to oxidative stress because the biosynthesis of these neurotransmitters from tyrosine requires molecular oxygen. Moreover, the autooxidization of catecholamines, the end product of which is melanin that accumulates in neurons of the substantia nigra and locus ceruleus, leads to the generation of ROS (H_2O_2, O_2^-, and OH·). The strong binding of DCD to these neurons is consistent with the roles of a 30-aa neural survival factor (21) and a cachexia factor possibly derived from DCD (20). Definition of the relationships among these peptides requires further studies.

In summary, DCD is a novel growth and survival factor that is overexpressed in ≈10% of primary invasive breast carcinomas, and its overexpression, at least in some cases, is associated with a gain of its locus at 12q13.1. Based on its function and restricted expression pattern in normal adult tissues, DCD is a candidate cancer therapeutic target. The secreted nature and extracellular mechanism of DCD action make it even more attractive for such a purpose.

Neurons are particularly sensitive to ROS, whereas tumor cells themselves produce large amounts of ROS (29). Therefore, the high expression of DCD in these cell types may be essential for their survival. Thus, therapeutic activation of the DCD-signaling pathway may be beneficial in certain neurodegenerative diseases involving catecholaminergic neurons such as Parkinson's disease, and its therapeutic inhibition may be an effective treatment of tumors with DCD expression.

We thank Drs. Massimo Loda, Marcus Bosenberg, and Gabriela Lodeiro for their help with this study. We are indebted to Wen-Lin Kuo, Fred Waldman, Sandy DeVries, and Joe W. Gray (University of California, San Francisco) for providing their BAC array CGH data before publication. This work was supported in part by the National Cancer Institute Cancer Genome Anatomy Project and Specialized Programs of Research Excellence in Breast Cancer at Dana–Farber/Harvard Cancer Center (to K.P.), Johns Hopkins University (to Y.K.B. and E.G.), the University of California, San Francisco (to Joe W. Gray), National Institute on Aging Grant AG05128 (to C.M.H.), and U.S. Army postdoctoral fellowships (to P.S., A.M.-S., and D.P.).

1. Greenlee, R. T., Murray, T., Bolden, S. & Wingo, P. A. (2000) *CA Cancer J. Clin.* **50,** 7–33.
2. Perou, C. M., Sorlie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., Rees, C. A., Pollack, J. R., Ross, D. T., Johnsen, H., Akslen, L. A., *et al.* (2000) *Nature* **406,** 747–752.
3. Sorlie, T., Perou, C. M., Tibshirani, R., Aas, T., Geisler, S., Johnsen, H., Hastie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., *et al.* (2001) *Proc. Natl. Acad. Sci. USA* **98,** 10869–10874.
4. van de Vijver, M. J., He, Y. D., van't Veer, L. J., Dai, H., Hart, A. A., Voskuil, D. W., Schreiber, G. J., Peterse, J. L., Roberts, C., Marton, M. J., *et al.* (2002) *N. Engl. J. Med.* **347,** 1999–2009.
5. van't Veer, L. J., Dai, H., van de Vijver, M. J., He, Y. D., Hart, A. A., Mao, M., Peterse, H. L., van der Kooy, K., Marton, M. J., Witteveen, A. T., *et al.* (2002) *Nature* **415,** 530–536.
6. Porter, D., Lahti-Domenici, J., Keshaviah, A., Bae, Y. K., Argani, P., Marks, J., Richardson, A., Cooper, A., Strausberg, R., Riggins, G. J., *et al.* (2003) *Mol. Cancer Res.* **1,** 362–375.
7. Velculescu, V. E., Zhang, L., Vogelstein, B. & Kinzler, K. W. (1995) *Science* **270,** 484–487.
8. Polyak, K. & Riggins, G. J. (2001) *J. Clin. Oncol.* **19,** 2948–2958.
9. Kononen, J., Bubendorf, L., Kallioniemi, A., Barlund, M., Schraml, P., Leighton, S., Torhorst, J., Mihatsch, M. J., Sauter, G. & Kallioniemi, O. P. (1998) *Nat. Med.* **4,** 844–847.
10. Hulette, C. M., Welsh-Bohmer, K. A., Crain, B., Szymanski, M. H., Sinclaire, N. O. & Roses, A. D. (1997) *Arch. Pathol. Lab. Med.* **121,** 615–618.
11. Krop, I. E., Sgroi, D., Porter, D. A., Lunetta, K. L., LeVangie, R., Seth, P., Kaelin, C. M., Rhei, E., Bosenberg, M., Schnitt, S., *et al.* (2001) *Proc. Natl. Acad. Sci. USA* **98,** 9796–9801.
12. St. Croix, B., Rago, C., Velculescu, V., Traverso, G., Romans, K. E., Montgomery, E., Lal, A., Riggins, G. J., Lengauer, C., Vogelstein, B. & Kinzler, K. W. (2000) *Science* **289,** 1197–1202.
13. Qian, Y., Fritzsch, B., Shirasawa, S., Chen, C. L., Choi, Y. & Ma, Q. (2001) *Genes Dev.* **15,** 2533–2545.
14. Kuchinka, B. D., Kalousek, D. K., Lomax, B. L., Harrison, K. J. & Barrett, I. J. (1995) *Mod. Pathol.* **8,** 183–186.
15. Pinkel, D., Segraves, R., Sudar, D., Clark, S., Poole, I., Kowbel, D., Collins, C., Kuo, W. L., Chen, C., Zhai, Y., *et al.* (1998) *Nat. Genet.* **20,** 207–211.
16. Flanagan, J. G. & Leder, P. (1990) *Cell* **63,** 185–194.
17. Lal, A., Lash, A. E., Altschul, S. F., Velculescu, V., Zhang, L., McLendon, R. E., Marra, M. A., Prange, C., Morin, P. J., Polyak, K., *et al.* (1999) *Cancer Res.* **59,** 5403–5407.
18. Porter, D. A., Krop, I. E., Nasser, S., Sgroi, D., Kaelin, C. M., Marks, J. R., Riggins, G. & Polyak, K. (2001) *Cancer Res.* **61,** 5697–5702.
19. Burge, C. & Karlin, S. (1997) *J. Mol. Biol.* **268,** 78–94.
20. Cariuk, P., Lorite, M. J., Todorov, P. T., Field, W. N., Wigmore, S. J. & Tisdale, M. J. (1997) *Br. J. Cancer* **76,** 606–613.
21. Cunningham, T. J., Hodge, L., Speicher, D., Reim, D., Tyler-Polsz, C., Levitt, P., Eagleson, K., Kennedy, S. & Wang, Y. (1998) *J. Neurosci.* **18,** 7047–7060.
22. Schittek, B., Hipfel, R., Sauer, B., Bauer, J., Kalbacher, H., Stevanovic, S., Schirle, M., Schroeder, K., Blin, N., Meier, F., *et al.* (2001) *Nat. Immunol.* **2,** 1133–1137.
23. Spitz, D. R., Sim, J. E., Ridnour, L. A., Galoforo, S. S. & Lee, Y. J. (2000) *Ann. N.Y. Acad. Sci.* **899,** 349–362.
24. Sanghi, S., Kumar, R., Lumsden, A., Dickinson, D., Klepeis, V., Trinkaus-Randall, V., Frierson, H. F., Jr., & Laurie, G. W. (2001) *J. Mol. Biol.* **310,** 127–139.
25. McDevitt, T. M. & Tisdale, M. J. (1992) *Br. J. Cancer* **66,** 815–820.
26. Todorov, P. T., McDevitt, T. M., Cariuk, P., Coles, B., Deacon, M. & Tisdale, M. J. (1996) *Cancer Res.* **56,** 1256–1261.
27. Todorov, P., Cariuk, P., McDevitt, T., Coles, B., Fearon, K. & Tisdale, M. (1996) *Nature* **379,** 739–742.
28. Cunningham, T. J., Jing, H., Akerblom, I., Morgan, R., Fisher, T. S. & Neveu, M. (2002) *Exp. Neurol.* **177,** 32–39.
29. Szatrowski, T. P. & Nathan, C. F. (1991) *Cancer Res.* **51,** 794–798.

ARTICLE

Molecular characterization of the tumor microenvironment in breast cancer

Minna Allinen,[1,3] Rameen Beroukhim,[1,3,10] Li Cai,[2,10] Cameron Brennan,[1,10] Jaana Lahti-Domenici,[1] Haiyan Huang,[2,6,11] Dale Porter,[1,3] Min Hu,[1,3] Lynda Chin,[1,4] Andrea Richardson,[5,7] Stuart Schnitt,[5,8] William R. Sellers,[1,3,9] and Kornelia Polyak[1,3]

[1]Department of Medical Oncology
[2]Department of Research Computing
Dana-Farber Cancer Institute, Boston, Massachusetts 02115
[3]Department of Medicine
[4]Deparment of Dermatology
[5]Department of Pathology
Harvard Medical School, Boston, Massachusetts 02115
[6]Department of Biostatistics, Harvard School of Public Health, Boston, Massachusetts 02115
[7]Department of Pathology, Brigham and Women's Hospital, Boston, Massachusetts 02115
[8]Department of Pathology, Beth Israel Deaconess Medical Center, Boston, Massachusetts 02115
[9]Broad Institute of Harvard and MIT, Cambridge, Massachusetts 02142
[10]These authors contributed equally to this work.
[11]Present address: Department of Statistics, University of California, Berkeley, Berkeley, California 94720.
*Correspondence: kornelia_polyak@dfci.harvard.edu

Summary

Here we describe the comprehensive gene expression profiles of each cell type composing normal breast tissue and in situ and invasive breast carcinomas using serial analysis of gene expression. Based on these data, we determined that extensive gene expression changes occur in all cell types during cancer progression and that a significant fraction of altered genes encode secreted proteins and receptors. Despite the dramatic gene expression changes in all cell types, genetic alterations were detected only in cancer epithelial cells. The CXCL14 and CXCL12 chemokines overexpressed in tumor myoepithelial cells and myofibroblasts, respectively, bind to receptors on epithelial cells and enhance their proliferation, migration, and invasion. Thus, chemokines may play a role in breast tumorigenesis by acting as paracrine factors.

Introduction

Breast cancer is the most commonly identified and one of the deadliest neoplasms in women in Western countries. The recent trend toward improvement in breast cancer mortality rate is largely due to increased diagnosis of early stage disease, while our therapeutic options for advanced stage breast carcinomas are still fairly limited. Thus, there is a need to better understand the molecular basis of breast cancer initiation and progression and to use this knowledge for the design of targeted, molecular-based therapies. In the past few years, newly developed technologies such as microarrays and SAGE (serial analysis of gene expression) have enabled us to analyze molecular differences between normal and cancer cells at a genome-wide level in comprehensive and unbiased ways (Schena et al., 1995; Velculescu et al., 1995). Using these approaches, the molecular-based classification of breast cancer has become a reality, and molecular signatures correlating with metastatic behavior and clinical outcome have been identified (Ramaswamy et al., 2003; Sorlie et al., 2001; van 't Veer et al., 2002; van de Vijver et al., 2002). However, since most of these analyses were performed using bulk tissue samples that are composed of multiple cell types or purified tumor epithelial cells, the specific contribution of epithelial and stromal cells to these tumor classifiers and prognostic signatures is unknown. Similarly, in the past decades the major focus of cancer research has been the transformed tumor cell itself, while the role of the cellular microenvironment in tumorigenesis has not been widely explored. Early studies demonstrated the ability of stromal tissues to regulate the growth and differentiation state of breast cancer cells (DeCosse

SIGNIFICANCE

Despite compelling cell biological studies and histopathological observations incriminating myoepithelial and stromal cells in tumorigenesis, our knowledge of the genes that mediate changes in the tumor microenvironment and interactions among various cell types in breast cancer and their role in tumorigenesis is limited. Similarly, the occurrence and role of genetic changes in stromal cells are undefined. Here, we describe a comprehensive molecular characterization of each cell type composing normal breast tissue and in situ and invasive breast carcinomas. We identified several genes as potential mediators of epithelial-stromal/myoepithelial cell interactions, including the CXCL12 and CXCL14 chemokines. These data should therefore provide a valuable resource for future basic and clinical studies addressing the role of epithelial-stromal/myoepithelial cell interactions in breast cancer.

et al., 1973, 1975), and several recent in vivo and in vitro studies have demonstrated that the growth, differentiation, invasive behavior, and polarity of normal mammary epithelial cells and breast carcinomas are influenced by surrounding stromal cells including fibroblasts, myofibroblasts, leukocytes, and myoepithelial cells (Bissell and Radisky, 2001; Radisky et al., 2001; Tlsty, 2001). In addition, certain histopathological features of breast tumors, including lymphocytic infiltration, fibrosis, and angio- and lymphangiogenesis, have proven prognostic significance. Despite these convincing data implicating a role for the tumor microenvironment in breast tumorigenesis, our understanding of the genes mediating cellular interactions and paracrine regulatory circuits among various cell types in normal and cancerous breast tissue and their role in breast tumorigenesis is limited.

In the past few years, the role of the cellular microenvironment in tumorigenesis has become an intense area of research. This is in part due to studies demonstrating that genetic abnormalities, such as loss of heterozygosity (LOH), occur not only in cancer cells, but in stromal cells as well (Kurose et al., 2001, 2002; Lakhani et al., 1998; Moinfar et al., 2000). However, no genes presumably targeted by these genetic events in stromal cells have been identified; thus, their role in breast tumorigenesis is still unknown.

As a consequence of studies focusing almost exclusively on cancer cells, nearly all of the currently used cancer therapeutic agents target the cancer cells that, due to their inherent genomic instability, frequently acquire therapeutic resistance (Rajagopalan et al., 2003). In part due to frequent therapeutic failures during the course of treatment of advanced stage tumors, increasing emphasis has been placed on targeting various stromal cells, particularly endothelial cells, via therapeutic interventions. Since these cells are thought to be normal and genetically stable, they are less likely to develop acquired resistance to cancer therapy. Thus, isolating and characterizing each cell type (epithelial, myoepithelial, and various stromal cells) comprising nonmalignant and cancerous breast tissue would not only help us to understand the role these cells play in breast tumorigenesis, but would likely give us new molecular targets for cancer intervention and treatment.

Results

Purification of all cell types present in breast tissue

To determine the molecular profile of each cell type that, together, compose the breast tissue and to identify autocrine and paracrine interactions that may play a role in breast tumor progression, we developed a purification procedure that allows the isolation of pure cell populations from normal breast tissue and from in situ (ductal carcinoma in situ, DCIS) and invasive breast carcinomas (Figure 1A). We utilized cell type-specific cell surface markers and magnetic beads for the rapid sequential isolation of the various cell types. We used the BerEP4 antigen restricted to epithelial cells, the CD45 panleukocyte marker, and the P1H12 antibody that specifically recognizes endothelial cells. The CD10 antigen is present in myoepithelial cells and myofibroblasts, but also in some leukocytes. Thus, to minimize the crosscontamination of these different cell types, in the case of normal (N-MYOEP-1) and DCIS breast tissue, myoepithelial cells were isolated from organoids (breast ducts), while in invasive tumors we first removed the leukocytes prior to capturing the myofibroblasts using CD10 beads. Several recent studies reported that some morphologically distinct myoepithelial cells lack CD10 and other myoepithelial cell markers (Zhang et al., 2003). Thus, due to the use of CD10 beads for the isolation of myoepithelial cells, a subset of myoepithelial cells may have been excluded from our study. We were not able to identify an antibody that would specifically recognize fibroblasts and allow their purification; thus, we used the unbound fraction following the removal of all other cell types as a fibroblast-enriched "stromal" fraction. A detailed description of the purification method is described in the Supplemental Data (http://www.cancercell.org/cgi/content/full/6/1/■■■/DC1). Since this protocol includes sequential enzymatic digestion of the tissue, the possibility that the expression of some genes could be altered due to the procedure cannot be excluded. However, since we were able to verify the SAGE data by alternative methods using unprocessed tissue (Figure 3), these changes (if any) are likely to be minimal. The success of the purification method and the purity of each cell fraction were confirmed by performing RT-PCR on a small fraction of the isolated cells using cell type-specific genes (Figure 1B). The remaining portion of the cells (~10,000–100,000 cells, depending on the sample) was used for the generation of micro-SAGE libraries following previously described protocols (Porter et al., 2001, 2003a) and for the isolation of genomic DNA to be used for array comparative genomic hybridization (aCGH) and single nucleotide polymorphism (SNP) array studies. We have generated SAGE libraries from epithelial and myoepithelial cells (myofibroblasts from invasive tumors), infiltrating lymphocytes, endothelial cells, and fibroblasts (stroma) from one normal breast reduction tissue, two different DCIS, and three invasive breast tumors. Not all libraries were generated from all cases due to our inability to obtain sufficient amounts of purified cells. In addition, we also included a fibroadenoma and a phyllodes tumor in our SAGE analyses. Fibroadenomas are the most common benign breast tumors that are not considered to progress to malignancy despite genetic changes detected in the stromal (but not epithelial) cells (Amiel et al., 2003). Phyllodes tumors, on the other hand, are rare fibroepithelial tumors that are usually benign but can recur and progress to malignant sarcomas. Initially, phyllodes tumors were considered stromal neoplasms, but recent molecular studies demonstrating (frequently discordant) genetic alterations in both epithelial and stromal cells suggests that phyllodes tumors may represent a true clonal co-evolution of malignant epithelial and stromal cells (Sawyer et al., 2000, 2002). A detailed description of the tissue samples and the SAGE libraries is included in the Supplemental Data online. Analysis of the SAGE data confirmed that the cell purification procedure worked well, since several genes known to be specific for a particular cell type were present in the appropriate SAGE libraries. For example, cytokeratins 8 and 19, E-cadherin, HIN-1, and CD24 were highly specific for epithelial cells (HIN-1 only for normal epithelial cells); myofibroblast and myoepithelial cells demonstrated high levels of smooth muscle actin, various extracellular matrix proteins including collagens, and matrix metalloproteinases; and leukocyte libraries had the highest levels of several chemokines and lysozyme (Table 1 and Supplemental Table S1). In general, SAGE libraries prepared from the same cell type purified from different tissue samples were highly similar to each other, although there were differences as well, likely due to variability among patients and also slight variability

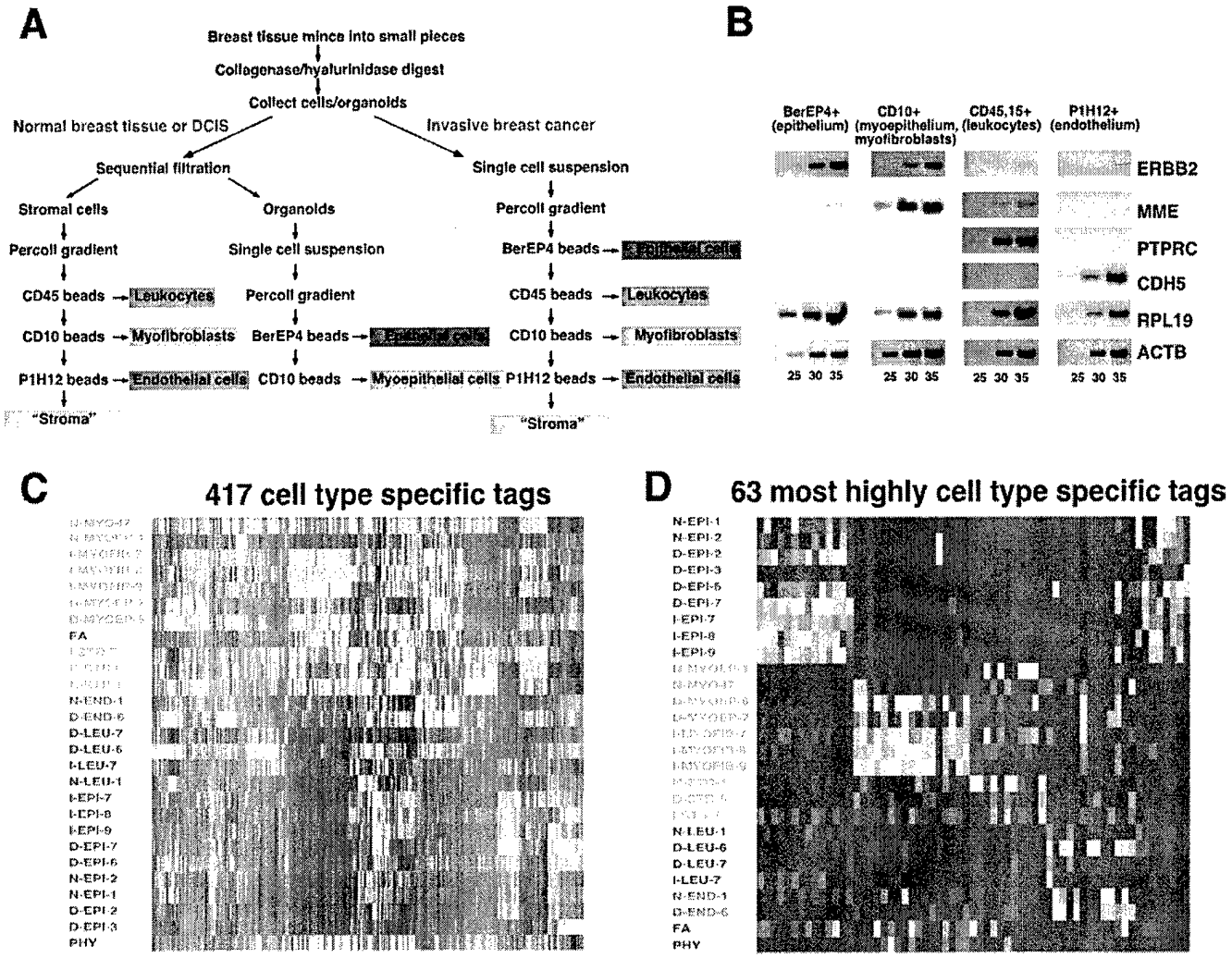


Figure 1. Isolation and characterization of each cell type comprising normal and cancerous breast tissue

A: Schematic outline of tissue fractionation and sequential purification of the various cell types from normal breast tissue and in situ and invasive breast carcinomas. The procedure is described in detail in the online Supplemental Data.

B: RT-PCR analysis of each cell fraction isolated from DCIS-7 using known cell type-specific genes to confirm the purity of the cells and integrity of the mRNA. MME (CD10) is highly specifically expressed in $CD10^+$ myoepithelial cells and myofibroblasts, PTPRC (CD45) in leukocytes, and CDH5 (endothelial cadherin) in endothelial cells. Although ERBB2 is not an absolutely epithelial cell-specific gene, its abundance is highest in luminal epithelial cells. PCR was performed at 25, 30, and 35 cycles. Genes expressed at equal levels in all cell types; β-actin (ACTB) and ribosomal protein L19 (RPL19) were used as controls.

C: Heat map depicting the relatedness of the different SAGE libraries based on 417 cell type-specific tags. Color scheme: blue, downregulated (low tag counts); green, mean tag counts; yellow, upregulated (high tag counts). The names of the SAGE libraries prepared from epithelial cells are in red, myoepithelial cells and myofibroblasts in green, stroma in yellow, leukocytes in blue, endothelial cells in pink, and fibroadenoma and phyllodes tumor (stroma fraction) in purple. A detailed description of the SAGE libraries and tissue samples is included in Supplemental Data.

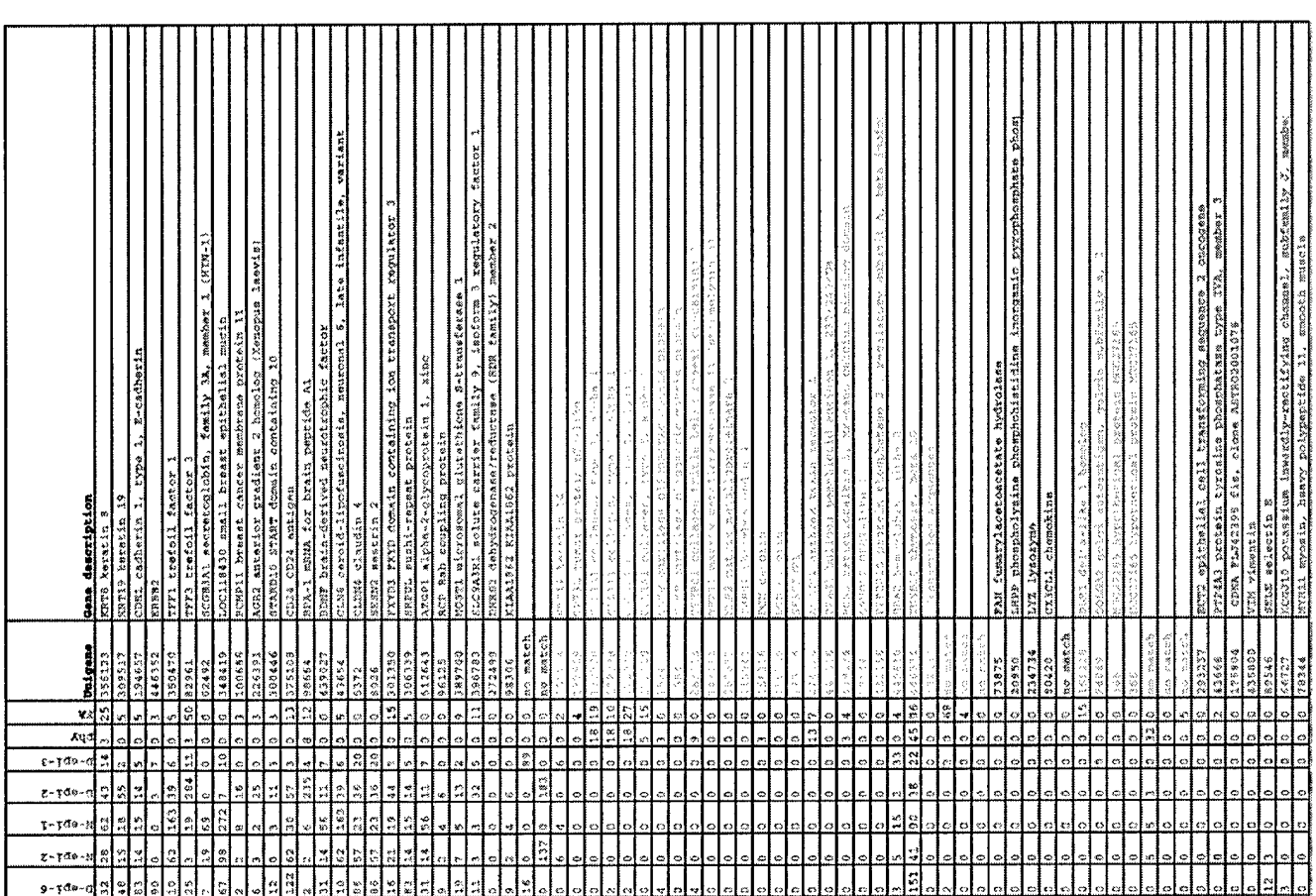
D: Heat map depicting the relatedness of the different SAGE libraries based on the 63 most highly cell type-specific tags. Color scheme and SAGE library names are described as above.

in the purification procedure itself (see Supplemental Data for more details).

Comprehensive gene expression profile of each cell type

Based on statistical methods developed for the analysis of SAGE data (see Experimental Procedures, Supplemental Data; Cai et al., 2004), we identified genes that are specifically expressed in a particular cell type and tumor progression stage (Tables 1 and 2 and Supplemental Tables S1–S15). Genes were defined as specific for a particular cell type if the average tag number in all the SAGE libraries generated from the selected cell type was statistically significantly ($p < 0.02$) different from all other cell types. For the purpose of these comparisons, we considered myoepithelial cells and myofibroblasts as one group due to their high degree of similarity, although there are genes that are specific for myoepithelial cells and myofibroblasts, respectively (Ronnov-Jessen et al., 1996). Using these criteria,

Table 1. List of 63 most highly cell type-specific tags and corresponding genes



SAGE tag numbers in the various libraries are indicated. Coloring reflects tag abundance in the different cell types.

Table 2. List of genes encoding secreted proteins and receptors overexpressed in DCIS myoepithelial cells compared to normal myoepithelium

SAGE Tag	N-MYOEP-1	D-MYOEP-7	D-MYOEP-6	Ratio D/N	Unigene	Gene description
ACCAAAAACC	2	274	849	244	172928	COL1A1 collagen, type I, alpha 1
GATCAGGCCA	0	191	181	124	443625	COL3A1 collagen, type III, alpha 1
TGGAAATGAC	0	50	228	93	172928	COL1A1 collagen, type I, alpha 1
CGGGGTGGCC	0	193	24	73	1584	COMP cartilage oligomeric matrix protein
CTAACGGGGC	0	169	20	63	513022	ISLR immunoglobulin superfamily containing leucine-rich repeat
CAGATAAGTT	0	72	101	58	222171	KIAA0182 KIAA0182 protein
CCGGGGGAGC	0	110	61	57	172928	COL1A1 collagen, type I, alpha 1
GTCAAAATTT	0	110	47	52	458354	THBS2 thrombospondin 2
GTGCTAAGCG	3	308	141	49	420269	COL6A2 collagen, type VI, alpha 2
GACTTTGGAA	0	36	110	49	172928	COL1A1 collagen, type I, alpha 1
CGCCGACGAT	0	100	32	44	287721	G1P3 interferon, alpha-inducible protein (clone IFI-6-16)
TTGGGATGGG	0	103	29	44	296941	HFL1 H factor (complement)-like 1
CATATCATTA	0	21	94	38	435795	IGFBP7 insulin-like growth factor binding protein 7
TCCAGGAAAC	0	72	39	37	11590	CTSF cathepsin F
GGCCCCTCAC	0	74	22	32	274313	IGFBP6 insulin-like growth factor binding protein 6
ACATTCCAAG	0	50	42	31	245188	TIMP3 tissue inhibitor of metalloproteinase 3
ATAAAAAGAA	0	19	73	31	83942	CTSK cathepsin K
GACCAGCAGA	0	43	48	30	172928	COL1A1 collagen, type I, alpha 1
ACTTATTATG	2	107	30	30	156316	DCN decorin
GTGCGCTGAG	0	33	52	28	274485	HLA-C major histocompatibility complex, class I, C
TGCGCTGGCC	0	67	18	28	289019	LTBP3 latent transforming growth factor beta binding protein 3
AGGCTCCTGG	3	217	31	27	24395	CXCL14 chemokine
CTCAACCCCC	2	105	19	27	162757	LRP1 low density lipoprotein-related protein 1
CAGCGGCGGG	0	57	13	23	2420	SOD3 superoxide dismutase 3, extracellular
GGCACCTCAG	2	36	65	22	512234	IL6 interleukin 6
GCCTGTCCCT	0	50	13	21	821	BGN biglycan
ATTTCTTCAA	0	19	44	21	31386	SFRP2 secreted frizzled-related protein 2
TCGAAGAACC	2	60	34	21	445570	CD63 CD63 antigen
ACATTCTTTT	0	17	44	20	389964	GPNMB glycoprotein (transmembrane)
CTGTCAGCGT	0	29	32	20	283713	CTHRC1 collagen triple helix repeat containing 1
CAGCTGGCCA	0	36	22	19	445240	FBLN1 fibulin 1
ACTGAAAGAA	3	124	50	19	458355	C1S complement component 1, s subcomponent
TTCTGTGCTG	3	105	40	16	376414	C1R complement component 1, r subcomponent
GGATGTGAAA	0	19	26	15	283477	CD99 CD99 antigen
ACTCAGCCCG	2	36	28	14	101382	TNFAIP2 tumor necrosis factor, alpha-induced protein 2
TTTCCCTCAA	2	21	42	14	75111	PRSS11 protease, serine, 11 (IGF binding)
CTAAAAAAAA	0	26	15	14	54457	CD81 CD81 antigen (target of antiproliferative antibody 1)
GGCCACGTAG	0	26	15	14	155597	DF D component of complement
AAGAAAGGAG	0	21	20	14	202097	PCOLCE procollagen C-endopeptidase enhancer
GGAGGAATTC	0	21	20	14	418123	CTSL cathepsin L
AGCCACCGCG	2	43	19	14	355874	RABL2B RAB, member of RAS oncogene family-like 2B
TGTAAACAAT	0	19	22	14	170040	PDGFRL platelet-derived growth factor receptor-like
ACCTTGAAGT	2	36	19	12	407546	TNFAIP6 tumor necrosis factor, alpha-induced protein 6
CATAAATGCG	0	21	13	12	436042	CXCL12 chemokine (stromal cell-derived factor 1)
TTGCTGACTT	12	122	279	11	415997	COL6A1 collagen, type VI, alpha 1
ATGGCAACAG	0	17	17	11	149609	ITGA5 integrin, alpha 5
CTCTCCAAAC	2	26	20	10	384598	SERPING1 serine proteinase inhibitor, clade G, member 1
TGCCTGCACC	5	76	46	9	304682	CST3 cystatin C
GGAAATGTCA	18	93	325	8	367877	MMP2 matrix metalloproteinase 2
CAGGTTTCAT	12	124	117	7	24395	CXCL14 chemokine
CCGTGACTCT	12	112	70	5	433622	FSTL1 follistatin-like 1

SAGE tag numbers reflect tag numbers normalized to the SAGE library with the highest tag number. Ratio was calculated as a ratio of the average tag numbers in the two DCIS myoepithelial libraries divided by the tag numbers in the normal myoepithelial library. Genes highlighted in red were selected for follow-up studies.

we identified 357 tags that differentiate epithelial cells from other cell types, 572 tags specifying myoepithelial cells and myofibroblasts, 502 tags discriminating leukocytes, 604 tags selecting stroma, and 124 tags discerning endothelial cells from other cells. To further define SAGE tags specific for each cell type, within each group of tags we selected the ones that were not only statistically significantly different, but also more abundant in the specific cell type. This led to the identification of 70 tags that were most abundant in epithelial cells, 117 tags present at highest levels in myoepithelial cells and myofibroblasts, 70 tags highly expressed in leukocytes, and 117 stroma- and 78 endothelium-specific tags (Supplemental Tables S3, S5, S7, S9, and S11). Several of these genes have previously been described as being specific for a particular cell type, such as keratins 8 and 19 for epithelial cells, keratins 14 and 17 for myoepithelial cells, and chemokines and chemokine receptors for leukocytes (Page et al., 1999), but the cell type-specific expression of the majority of the genes has not been documented. The majority of the transcripts corresponding to these cell type-specific SAGE tags encode known genes, but a significant fraction are uncharacterized ESTs or currently have no cDNA match (~10% of the tags on average belong to each of these last two groups). The only exceptions were tags most abundant in stroma, since in this group 25/117 tags (21%) had no database match, suggesting that they correspond to previously unidentified transcripts.

Next, using the SAGE tags most abundant in (417 tags) or most highly specific for (63 tags) each of the five cell types, we performed clustering analysis of all 27 SAGE libraries using a new Poisson model-based K-means algorithm (PK algorithm, Supplemental Data; Cai et al., 2004) to delineate similarities and differences among the samples (Figures 1C and 1D). In addition, we also performed clustering analysis of the SAGE libraries using each of the cell type-specific gene sets (Supplemental Figures S1 and S2). The PK clustering method orders the samples according to their relatedness. For example, using the 63 most highly cell type-specific SAGE tags, we obtained a division of the 27 SAGE libraries according to cell types, and within each cell type subgroup, the DCIS samples were located between normal breast tissue and invasive breast cancer SAGE libraries (Figure 1D). This result indicates that not only tumor epithelial cells, but also other cell types in the tumor, are different from their corresponding normal counterparts. Since these differences are already pronounced at a pre-invasive (DCIS) tumor stage, they suggest a role for stromal changes not only in tumor invasion and metastasis, but also in the earlier steps of breast tumorigenesis.

Based on our SAGE data, we found that the most consistent and dramatic gene expression changes occur in myoepithelial cells. More than 300 genes were differentially expressed at $p < 0.002$ in both DCIS myoepithelial libraries, and interestingly, a significant fraction of these genes (89 out of 245) encode secreted or cell surface proteins, suggesting extensive abnormal paracrine interactions between myoepithelial and other cell types (Supplemental Table S5). Myoepithelial cells are thought to be derived from bipotential stem cells that also give rise to luminal epithelial cells, although recently another progenitor has been identified that can differentiate only into myoepithelial cells (Bocker et al., 2002; Dontu et al., 2003). The function of myoepithelial cells and their role in breast cancer are not well understood, but myoepithelial cells have been shown to be able to suppress breast cancer cell growth, invasion, and angiogenesis (Deugnier et al., 2002; Sternlicht and Barsky, 1997). The main distinguishing feature of in situ and invasive carcinomas, which is also used as a diagnostic criteria, is that in DCIS, the cancer epithelial cells are separated from the stroma by a nearly continuous layer of myoepithelial cells and basement membrane, while in invasive and metastatic tumors, cancer cells are admixed with stroma. Due to our SAGE and previously published data suggesting a role for these cells in breast tumor progression, we focused our follow-up studies on myoepithelial cells with special emphasis on secreted proteins and receptors abnormally expressed in these cells. Several proteases (cathepsins F, K, and L, MMP2, and PRSS11), protease inhibitors (thrombospondin 2, SERPING1, cystatin C, and TIMP3), and many different collagens were highly upregulated in DCIS myoepithelial cells, suggesting a role for these cells in extracellular matrix remodeling (Table 2).

Analysis of the genotype of epithelial, myoepithelial, and stromal cells

To determine if the dramatic gene expression changes observed in tumor myoepithelial and stromal cell types could be due to underlying genetic alterations, we first performed aCGH analysis of epithelial and myoepithelial cells and of myofibroblasts from two DCIS (DCIS-6 and -7) and one invasive breast carcinoma (IDC7) used for SAGE. As expected, we detected numerous chromosomal gains and losses in the tumor epithelial cells, while no changes were detected in myoepithelial cells and myofibroblasts (Figures 2A and 2B). Similarly, no genetic changes were detected in epithelial and myoepithelial cells isolated from normal tissue adjacent to the tumors (Figure 2A). These data suggest that although nonepithelial cells in breast tumors are phenotypically distinct from their normal counterparts, genetic changes detectable by aCGH appear to be limited to cancer epithelial cells. However, since array CGH is thought to be more sensitive for the detection of copy number gains than losses and previous studies demonstrated LOH in stromal cells, we applied another technology, SNP arrays, for the analysis of isolated epithelial and stromal cells from a set of breast tumors. As expected, cancer epithelial cells from all but one invasive breast tumor demonstrated numerous LOH on nearly all chromosomes, while myofibroblasts and other stromal cells from the same tumors appeared to be mostly normal (Figure 2C and Supplemental Figure S3). Clustering analysis based on the inferred LOH data clearly divided the samples into two major groups, the tumor epithelial and stromal cells from different cases demonstrating more similarity to each other than to their corresponding other cell type (Figure 2C). The only exception was epithelial cells from IDC10 (a low-grade estrogen receptor-positive tumor) that did not appear to have major genetic changes (the purity of the tumor epithelial cells was confirmed by RT-PCR, data not shown), while in the phyllodes tumor, the stroma had numerous genetic alterations with much fewer LOH events detected in the epithelial cells. We did not detect significant LOH in the three fibroadenomas analyzed or in the one LCIS (lobular carcinoma in situ) case. Two nonepithelial samples (I-MYOFIB-8 and I-STR-13) had a few areas where 2–5 adjacent SNPs exhibited LOH (Figure 1C), but careful examination of these SNPs individually suggested that these LOH calls are likely due to poor hybridization results. In order to resolve this issue, we amplified and sequenced eight of these ambiguous

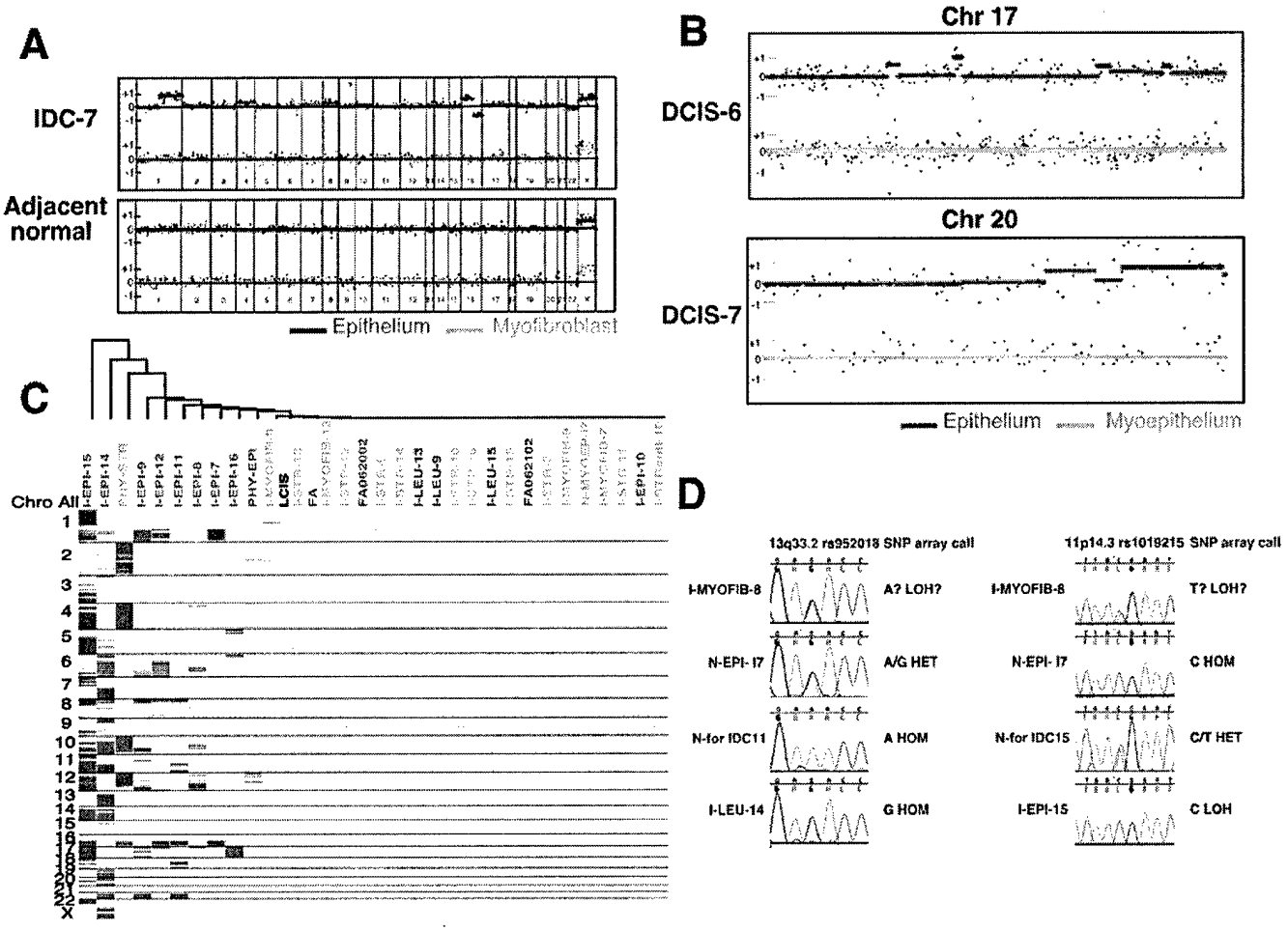


Figure 2. Genotype analysis of fractionated normal and tumor breast tissue

A: Array CGH analysis of luminal epithelial (red line) and myofibroblasts (green line) cells isolated from IDC-7 invasive breast tumor used for SAGE and from adjacent normal tissue. Mode centered segmented data, significant gains and losses defined as Log2 signal ratio of greater than or equal to +0.13 or −0.13, respectively, are depicted.

B: Array CGH analysis of luminal epithelial (red line) and myoepithelial (green line) cells from DCIS-6 and DCIS-7. Areas with statistically significant gains in the epithelial cells (chromosome 17 in the case of DCIS-6 and chromosome 20 for DCIS-7) are depicted, indicating that myoepithelial cells do not share these changes with the epithelial cells. No significant gains and losses were detected in any other areas of the genome in the myoepithelial cells (data not shown).

C: SNP array analysis of purified epithelial and stromal cells from invasive breast carcinomas, phyllodes tumor, fibroadenomas, and LCIS. Samples are clustered based on inferred loss of heterozygosity (LOH). All but one tumor epithelial DNA sample are clustered together to the left, while all stromal samples, regardless of their origin, are clustered together to the right. Inferred loss of heterozygosity (LOH) is indicated in blue, yellow indicates regions retaining heterozygosity, and white regions are indeterminate (noninformative). The names of DNA samples obtained from epithelial cells are depicted in red, myofibroblasts in green, stroma in yellow, leukocytes in blue, endothelial cells in pink, fibroadenoma in purple, and LCIS in black. A detailed description of the samples is included in the online Supplemental Data.

D: Sequence analysis of two ambiguous SNP cells present in I-MYOFIB-8 and in several controls. For all cases, the chromatograms of the sequence reads and the SNP array calls are indicated. One of the SNPs (rs952018) is on chromosome 13q33.2, while the other one (rs1019215) is on chromosome 11p14.3. As depicted in the figure in the case of SNP rs952018, the I-MYOFIB-8 sample had both "A" and "G" peaks just like the N-EPI-17 sample, proving the retention of both alleles, while the I-LEU-14 sample was homozygous for the "G" allele and the N-for IDC11 (normal DNA corresponding to tumor IDC11) was homozygous for the "A" allele. Similarly in the case of SNP rs952018, the I-MYOFIB-8 sample had both T and C peaks just like the N-for IDC15 (normal DNA corresponding to tumor IDC15), while the N-EPI-17 and I-EPI-15 were both homozygous for the "C" allele.

SNPs from these two stromal samples (I-MYOFIB-8 and I-STR-13) together with several controls, where the SNP results clearly depicted heterozygous or homozygous alleles. In all seven cases in which high-quality sequencing results were obtained, we found no evidence of LOH in either of these two ambiguous stromal samples (Figure 2D).

Evaluation of gene expression by immunohistochemistry and mRNA in situ hybridization

The generation of the SAGE libraries involved the in vitro purification of the cells that could potentially alter the in vivo gene expression patterns, although prior SAGE data from several

laboratories suggest that these changes are likely to be minimal (Porter et al., 2003a, 2003b; St Croix et al., 2000). However, in order to further investigate the expression of selected genes at the cellular level in vivo, we performed immunohistochemical analyses and mRNA in situ hybridization in a panel of DCIS and invasive breast tumors (including tumors used for SAGE as well as additional cases). In addition, the cell type specificity of some genes was verified by RT-PCR in the samples used for SAGE (data not shown). Immunohistochemical analysis confirmed that two genes, IL-1β and CCL3 (MIP1α), are highly expressed in leukocytes infiltrating DCIS, but not normal breast tissue, whereas the PTPRC (CD45) panleukocyte marker was expressed in both cases (Figure 3A). Despite the similar number of total leukocytes in invasive tumors, the frequency of IL-1β and CCL3-positive leukocytes was much lower than in DCIS, suggesting that in situ and invasive breast carcinomas may be immunologically dissimilar. mRNA in situ hybridization determined that in DCIS tumors, the expression of PDGF receptor β-like (PDGFRBL), cathepsin K (CTSK), and CXCL12 was localized to myofibroblasts as determined by smooth muscle actin (ACTA2) staining, CXCL14 was expressed only in myoepithelial cells, while TIMP3, cystatin C (CST3), and collagen triple helix repeat containing 1 (CTHRC1) were expressed in both myoepithelial cells and myofibroblasts. In invasive tumors, all seven genes were expressed in myofibroblasts. No signal was detected in normal breast tissue nor with the sense probes (Figure 3B, Supplemental Figure S4, and data not shown). Interestingly, although in DCIS tumors we detected CXCL14 expression only in myoepithelial cells, in some (4/9) invasive breast carcinomas, the expression of CXCL14 was restricted to the tumor epithelial cells (Figures 3B and 4A). Similarly, some breast cancer cell lines expressed high levels of CXCL12 or CXCL14 in vitro, suggesting that during tumor progression a paracrine factor may be converted into an autocrine one due to its upregulation in the tumor epithelial cells (Figure 4B). Interestingly, all CXCL14-positive invasive ductal carcinomas and even the CXCL14 expressing breast cancer cell line (UACC812) were obtained from young, premenopausal patients (average age of onset 39 years), suggesting a possible association of CXCL14 expression with hormone levels or clinico-pathologic characteristics of the tumors, the analysis of which requires the examination of larger tumor sets.

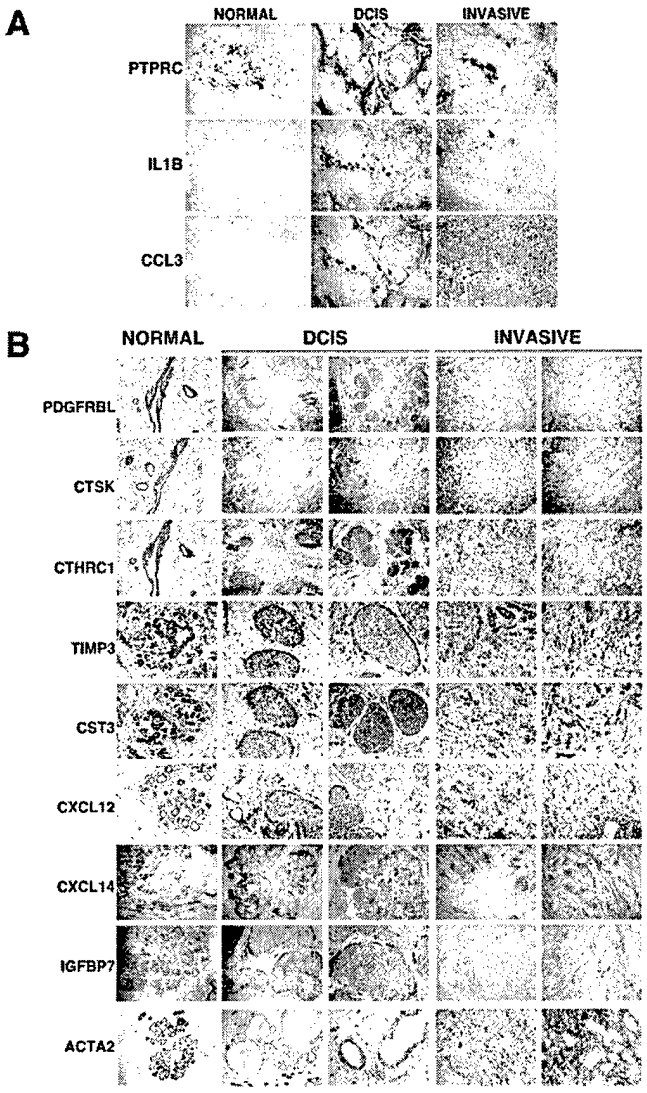


Figure 3. Validation of SAGE data using immunohistochemistry and mRNA in situ hybridization and Northern blot analysis

A: Immunohistochemical analysis of PTPRC (CD45), IL1β, and CCL3 expression in normal, DCIS, and invasive cancer breast tissue. Black signal indicates expression of the indicated proteins in leukocytes. Methyl green was used to stain the nuclei to visualize tissue histology. Magnification is 100×.

B: mRNA in situ hybridization analysis of the indicated genes using antisense ribo-probes in a panel of normal, DCIS, and invasive breast cancer tissue. Red (PDGFRL, CTSK, CTHRC1, TIMP3, CST3, and CXCL12) or black (CXCL14 and IGFBP7) staining indicates the presence of the mRNA depending on the hybridization protocol used. Paraffin sections were analyzed for ACTA2 (smooth muscle actin) expression by immunohistochemistry to confirm the identity of myoepithelial cells and myofibroblasts. Brown staining indicates the expression of SMA in myoepithelial cells and myofibroblasts. Magnification is 100×. More detailed images with higher (200×) magnification are included in Supplemental Data (Supplemental Figure S4).

The effect of CXCL12 and CXCL14 chemokines on breast cancer cells

The high level of expression of two chemokines, CXCL12 and CXCL14, in myoepithelial cells and myofibroblasts both in DCIS and invasive breast carcinomas was particularly interesting due to the known function of chemokines as regulators of cell proliferation, differentiation, migration, and invasion (Gerard and Rollins, 2001; Muller et al., 2001; Rossi and Zlotnik, 2000). To determine if CXCL12 and CXCL14 may act as autocrine and/or paracrine factors in breast tumors, we investigated which cell types appear to have receptors for these chemokines in vivo in primary breast tissue. The signaling receptor for CXCL12 is CXCR4, which is known to be widely expressed in various lymphoid as well as a variety of epithelial cells (Gerard and Rollins, 2001). We confirmed the expression of CXCR4 in lymphoid and breast epithelial cells using immunohistochemistry, while SAGE data indicated that its expression is increased in invasive tumors compared to DCIS and normal breast tissue (data not shown). The signaling receptor of CXCL14 is unknown, but cell surface ligand binding experiments have suggested the presence of a putative CXCL14 receptor on monocytes and B cells, suggesting that its receptor is not likely to be CXCR4 (Kurth et al., 2001; Sleeman et al., 2000). To determine if a

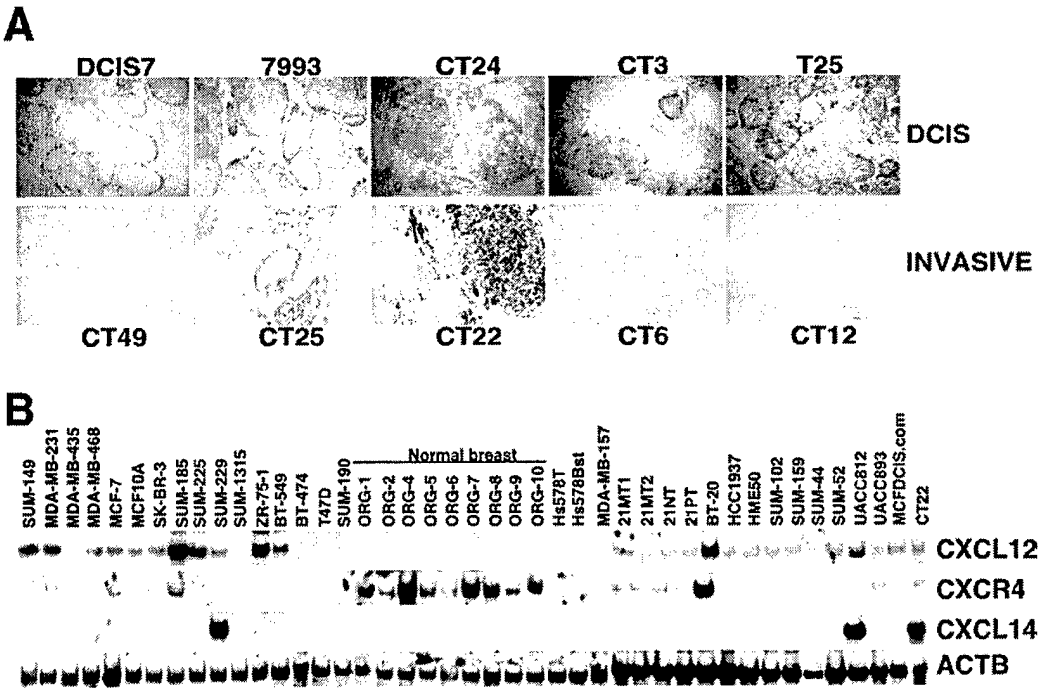


Figure 4. CXCL14 expression in primary breast tumors and breast cancer cell lines

A: mRNA in situ hybridization using CXCL14 antisense ribo-probe in multiple DCIS and invasive breast carcinomas including the tumors used for SAGE (DCIS-7). Black/purple staining indicates the presence of the CXCL14 mRNA, while nuclei were stained with nuclear FastRed to visualize tissue histology. The names of the tumor samples are indicated above/below of the pictures. In DCIS cases, CXCL14 is expressed only in myoepithelial cells, while in some invasive breast carcinomas (CT22 and CT25), strong expression is observed in tumor epithelial cells.

B: Northern blot analysis of CXCL12, CXCL14, and CXCR4 expression in the indicated breast cancer cell lines, breast organoids (ORG1–10, uncultured breast ducts from normal breast tissue), and primary breast tumor CT22. Hybridization with β-actin (ACTB) was used as a control for loading. Confirming the mRNA in situ hybridization data, strong CXCL14 expression is detected in tumor CT22, similarly in SUM-229 and UACC812 breast cancer cell lines.

CXCL14 binding cell surface protein(s) is also present on breast cancer cells, we generated an alkaline phosphatase-CXCL14 (AP-CXCL14) fusion protein to be used as a ligand in receptor binding assays. Conditioned media of AP-CXCL14 or control AP expressing cells was then used as an affinity reagent to stain normal and cancerous mammary tissue sections including the DCIS tumors used for SAGE. Blue staining indicated the presence of a CXCL14 binding protein in certain leukocytes and breast epithelial cells (Figure 5A). These results suggest the presence of a cell surface CXCL14 binding protein(s) in cancerous and normal mammary epithelial cells and are consistent with a paracrine mechanism of CXCL14 action in the breast. To test further the binding characteristics of AP-CXCL14, we performed in vitro ligand binding assays using various cell lines. Low-level AP-CXCL14 binding was detected in all cell lines tested, including MDA-MB-231 and MDA-MB-435 breast cancer and MCF10A immortalized mammary epithelial cells (data not shown). To further characterize the AP-CXCL14-putative CXCL14 receptor interaction, we performed more detailed binding assays in MDA-MB-231 breast cancer cells. Scatchard plot analysis showed two binding slopes in MDA-MB-231 cells indicating the presence of high-affinity ($K_d = 6.1 \times 10^{-8}$ M) and low-affinity ($K_d = 56.7 \times 10^{-8}$ M) binding sites (Figure 5B).

In previous studies, CXCL12 was demonstrated to enhance breast cancer cell growth, migration, and invasion (Hall and Korach, 2003; Muller et al., 2001). In order to determine if CXCL14 has similar effects, we tested the effect of conditioned medium containing AP-CXCL14 on the growth of MDA-MB-231 and MCF10A cells, while its effect on cell migration and invasion was investigated in MDA-MB-231 cells. Conditioned medias of cells transfected with AP alone and CXCL12 were used as negative and positive controls, respectively. Similar to CXCL12, CXCL14 enhanced the proliferation of MDA-MB-231 and MCF10A cells and the migration and invasion of MDA-MB-231 cells (Figures 5C and 5D, and data not shown). The concentration of AP-CXCL14 was 2–30 nM in these experiments, which is similar to the concentration required by several chemokines, including CXCL12, to exert biological effects. The same results were obtained in cell migration and invasion assays using CXCL14-AP (C-terminal AP-tag) and CXCL14-HA (C-terminal HA-tag) fusion proteins (Figure 5D and data not shown); thus, the observed effects are not likely to be due to the position or identity of the epitope tag. Preliminary results using recombinant CXCL14 protein and CXCL14 expressing adenovirus demonstrated possible induction of calcium flux in MDA-MB-231 and activation of AKT in MCF10A cells, respectively (data not shown), further suggesting that mammary epithelial cells have a functional CXCL14 receptor.

To determine if paracrine factors, including CXCL14, secreted by DCIS myoepithelial cells may influence the proliferation of tumor epithelial cells in vivo, we analyzed the expression of Ki67, a marker of cell proliferation, in the two DCIS samples

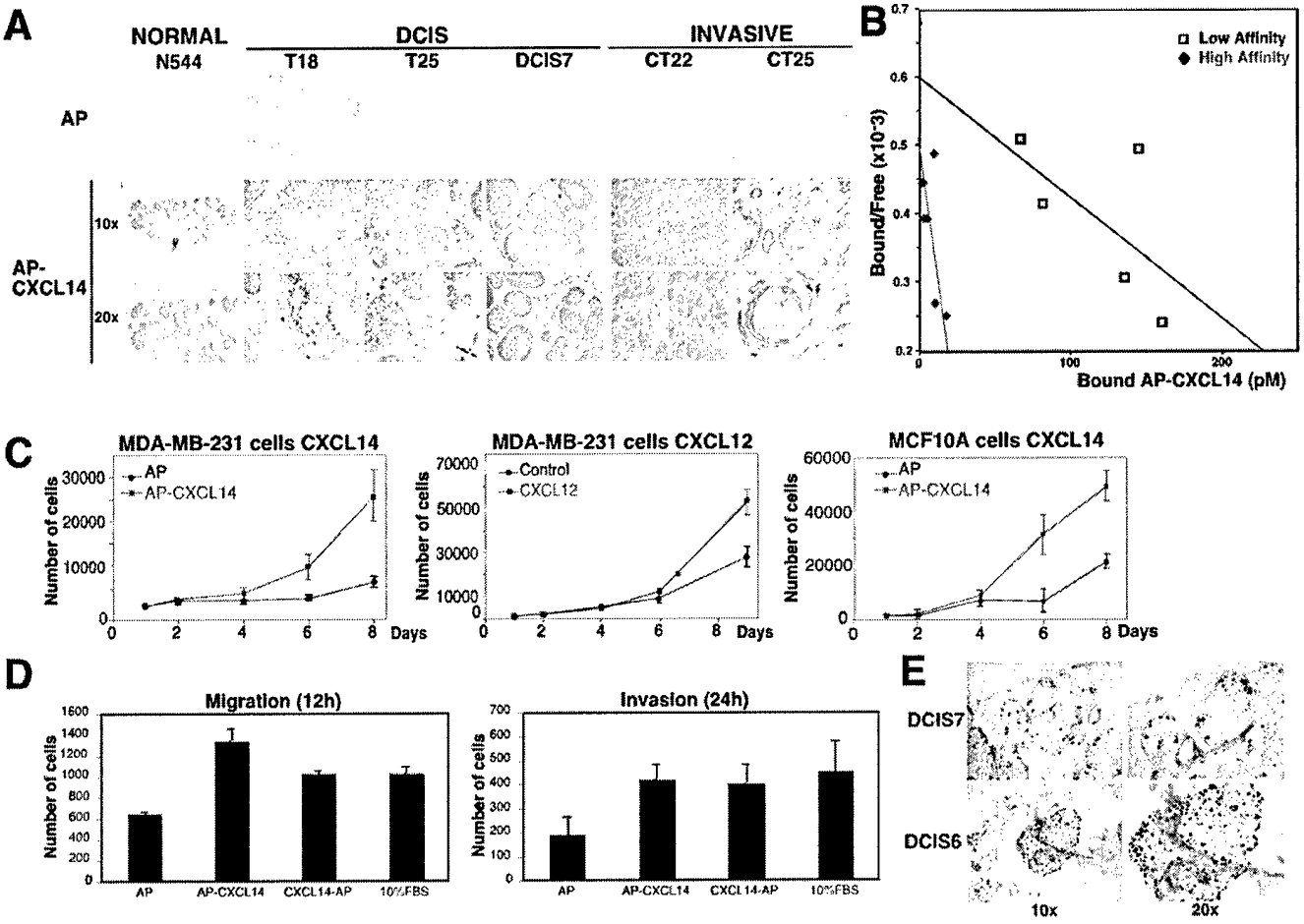


Figure 5. Analysis of CXCL14 ligand binding characteristics and function

A: Identification of a putative CXCL14 receptor in breast epithelial cells using an AP (alkaline phosphatase)-CXCL14 fusion protein as ligand. Blue staining reflecting AP activity indicates binding of AP-CXCL14 to breast epithelial cells and some stromal leukocytes, while no staining is detected with the AP alone negative control. All these tumor samples were also analyzed for the expression of CXCL14 by mRNA in situ hybridization (Figures 3B and 4A) and were expressing CXCL14 in tumor epithelial cells (CT22 and CT25) and DCIS myoepithelial cells (T18 and T25). Images were taken with 10× and 20× objectives (100× and 200× magnification).

B: Scatchard transformation of AP-CXCL14 binding assays in MDA-MB-231 cells. Red and black colored lines indicate high (K_d = 6.1 × 10^{-8} M) and low (K_d = 56.7 × 10^{-8} M) -affinity binding slopes, respectively.

C: The effect of CXCL12 and AP-CXCL14 on the growth of MDA-MB-231 breast cancer and MCF10A immortalized breast epithelial cells (red lines) compared to AP and control media (black lines). Representative result of experiments performed in triplicate.

D: The effect of CXCL12, CXCL14, and 10% fetal bovine serum (FBS) on the migration and invasion of MDA-MB-231 breast cancer cells. The number of cells that crossed the uncoated (migration) or Matrigel-coated membranes (invasion) is indicated on the y axis. Representative result of experiments performed in triplicate.

E: Immunohistochemical analysis of Ki67 expression in DCIS-6 and DCIS-7 samples to identify proliferating cells. Images were taken with 10× and 20× objectives (100× and 200× magnification). Ki67 is expressed in all phases of the cell cycle except in noncycling (G_0) cells. Tumor epithelial cells adjacent to the myoepithelial cell layer are more frequently positive than their more centrally located counterparts.

used for SAGE (Figure 5E). In both cases, epithelial cells adjacent to the myoepithelial cell layer were more frequently positive for Ki67 than tumor epithelial cells in other parts of the ducts. This result suggests that tumor epithelial cells may receive paracrine signals from adjacent myoepithelial cells that enhance their proliferation, although other reasons for this intraductal location-dependent proliferation difference cannot be excluded. Correlating with this, a recent study described that the gene expression profile of tumor epithelial cells located at the periphery and the center of DCIS ducts is significantly different (Zhu et al., 2003).

Discussion

Epithelial-mesenchymal interactions are known to be important for the normal development of the mammary gland and to play a role in breast tumorigenesis (Bissell et al., 2002; Coussens and Werb, 2002; Kenny and Bissell, 2003; Radisky et al., 2001; Shekhar et al., 2003; Tlsty, 2001; Tlsty and Hein, 2001; Wiseman and Werb, 2002). Early studies demonstrated that the normal mammary microenvironment is capable of "reverting" the neoplastic phenotype of breast cancer cells by inducing cellular differentiation (DeCosse et al., 1973, 1975), suggesting that

cancer cells can thrive only in a distorted environment or have to become independent of extracellular signals. The contribution of genetic host factors to tumor initiation, progression, and angiogenesis also support a role for nonepithelial cells in carcinogenesis (Hunter, 2004; Rohan et al., 2000). This was dramatically illustrated by the finding that inactivation of TGF-β type II receptor in stromal fibroblasts led to prostate and gastric epithelial neoplasia (Bhowmick et al., 2004). Similarly, a recent finding demonstrating that mammary tumors only formed in cleared mammary fat pads of rats treated with carcinogen, regardless of whether the injected epithelial cells were treated with carcinogen in vitro, also emphasizes the importance of stromal alterations in the initiating steps of breast cancer (Maffini et al., 2004). Numerous in vitro and in vivo studies using diverse experimental systems have demonstrated that the growth, survival, polarity, and invasive behavior of breast cancer cells can be modulated by myoepithelial and various stromal cells, and several genes have been implicated to play an important role in this process (Bissell et al., 2002; Coussens and Werb, 2002; Deugnier et al., 2002; Elenbaas and Weinberg, 2001; Gudjonsson et al., 2002; Kenny and Bissell, 2003; Radisky et al., 2001; Shekhar et al., 2003; Sternlicht and Barsky, 1997; Tlsty, 2001; Tlsty and Hein, 2001; Wiseman and Werb, 2002). However, comprehensive molecular analysis of all cell types that compose normal human mammary breast tissue and breast carcinomas has not been performed.

With the aim of delineating epithelial-stromal/myoepithelial cell interactions at the molecular level, we determined the comprehensive gene expression and genomic profiles of epithelial, myoepithelial, and stromal cells in normal breast tissue and in situ and invasive breast carcinomas. Our results confirm at the molecular level that the cellular microenvironment is dramatically different between normal breast tissue and breast carcinomas and that this is already evident at the in situ carcinoma stage. Based on our gene expression data, we determined that the most dramatic and consistent changes occur in myoepithelial cells and myofibroblasts and the majority of the differentially expressed genes encode secreted and cell surface proteins (Tables 1 and 2 and Supplemental Tables S2 and S5). Since previous data also implicated these two cell types in breast tumor progression, particularly in the transition of in situ to invasive carcinomas (Alpaugh et al., 2000; Barsky, 2003; Chauhan et al., 2003; Nguyen et al., 2000; Shao et al., 1998; Sternlicht and Barsky, 1997; Sternlicht et al., 1997; Walter-Yohrling et al., 2003), we mainly focused on tumor myoepithelial cells and myofibroblasts and the genes expressed by them.

Myoepithelial cells play a major role in the formation of the basement membrane and lactation due to their expression of type IV collagen, laminin, smooth muscle actin, and oxytocin receptor (Gudjonsson et al., 2002; Murrell, 1995). They also have been suggested to suppress breast cancer cell growth, invasion, and angiogenesis via shedding of CD44 and expression of protease inhibitors (Alpaugh et al., 2000; Barsky, 2003; Xiao et al., 1999). On the other hand, myoepithelial cells are also important for the survival, differentiation, and polarity of normal luminal epithelial cells (Gomm et al., 1997a, 1997b). Proteomic and mRNA expression profiling of short-term cultured myoepithelial cells and myoepithelial cell lines, respectively, gave a glimpse of the molecular basis for the tumor and invasion suppressor role of normal myoepithelium (Barsky, 2003; Page et al., 1999). Our SAGE-based profiling of freshly isolated, uncultured myoepithelial cells from normal breast tissue also demonstrated the high expression of laminin, tenascin, thrombospondin, and PAI-1 binding protein. However, the expression of these genes was downregulated in DCIS myoepithelial cells similar to that of cytokeratins 7, 14, and 17, oxytocin receptor, and tropomyosin, suggesting that DCIS myoepithelial cells are phenotypically altered and less differentiated than normal myoepithelial cells. Keeping with this, several recent studies described a lack of commonly used myoepithelial markers (including CD10 and SMA) in a subset of morphologically distinct myoepithelial cells, suggesting that myoepithelial cells may also be subject to pathological alterations (Zhang et al., 2003). Moreover, in support of a role for myoepithelial cells in breast tumor progression, it was recently reported that DCIS tumor epithelial cells adjacent to a disrupted myoepithelial cell layer are molecularly and genetically different from their more distant counterparts (Man et al., 2003).

Myofibroblasts are stromal fibroblasts with features of both myoblasts (e.g., expression of smooth muscle actin) and fibroblasts that have been implicated in breast cancer invasion, extracellular matrix remodeling, wound healing, and chronic inflammation (De Wever and Mareel, 2003; Gabbiani, 1999; Schurch, 1999). The cell type of origin of myofibroblasts is not conclusively established. Certain cytokines can induce (TGF-β) or inhibit (IFN-γ) the transformation of fibroblasts into myofibroblasts in vitro (De Wever and Mareel, 2003; Tanaka et al., 2003), while PDGF-B stimulates the proliferation of fibrocytes and their conversion into myofibroblasts in vivo (Oh et al., 1998). Isolation of various stromal and epithelial cells from breast tumors and their coculturing in vitro demonstrated that cancer epithelial cells can induce the expression of myofibroblast markers in a subset of fibroblasts (Ronnov-Jessen et al., 1995). However, the finding that only a small fraction of fibroblasts were transformed into myofibroblasts (Ronnov-Jessen et al., 1995) raises the question of whether myofibroblasts could be derived from specific stem cells that are normally present in the breast or in the bone marrow and are growth stimulated or recruited by adjacent cancer epithelial cells. Recent data both in animal models and human breast tumors support the hypothesis that at least a subset of cancer-associated myofibroblasts is derived from circulating bone-marrow derived cells (Chauhan et al., 2003; Ishii et al., 2003). Our finding that the gene expression profiles of myofibroblasts isolated from different invasive breast carcinomas are highly similar also suggest their common cell type of origin.

Two genes highly expressed in tumor myoepithelial cells and myofibroblasts encoding chemokines CXCL12 and CXCL14 were particularly interesting due to the recently demonstrated role of chemokines and chemokine receptors in cancer cell growth, invasion, and metastasis (Barbero et al., 2002; Chen et al., 2003; Hall and Korach, 2003; Kang et al., 2003; Muller et al., 2001; Scotton et al., 2002). CXCL12 has been previously implicated in breast cancer metastasis (Kang et al., 2003; Muller et al., 2001), but its high expression in DCIS (a pre-invasive tumor) myofibroblasts suggests that it might have additional roles in the earlier stages of breast tumorigenesis. Correlating with this hypothesis, CXCL12 was recently identified as a transcriptional target of the estrogen receptor that mediates estrogen-induced proliferation of breast cancer cells (Hall and Korach, 2003). Relatively little is known about the CXCL14 chemokine despite the fact that it was independently identified

by multiple labs using different approaches. The high expression of CXCL14 in inflammatory cells in multiple cancer types and its selectivity to monoctyes may suggest a role in macrophage development (Frederick et al., 2000; Hromas et al., 1999; Kurth et al., 2001; Sleeman et al., 2000). Although the receptor for CXCL14 has not been identified, the induction of calcium mobilization by recombinant CXCL14 in monocytes suggests that similar to other chemokines, it is also likely to signal via a G protein-coupled receptor. Our preliminary results demonstrating intracellular calcium flux in MDA-MB-231 breast cancer cells also support this hypothesis.

In addition to phenotypic alterations, several recent studies described genetic changes (including LOH and mutations in tumor suppressor genes) in stromal cells adjacent to breast cancer cells (Kurose et al., 2001, 2002; Lakhani et al., 1998; Moinfar et al., 2000; Wernert et al., 2001). Loss of heterozygosity at several loci has also been demonstrated in normal-appearing epithelial cells adjacent to breast carcinomas and short-term cultured luminal and myoepithelial cells (Deng et al., 1996; Forsti et al., 2001; Lakhani et al., 1999; Moinfar et al., 2000). In several cases, the tumor epithelial and stromal cells had discordant genetic changes, suggesting a clonal co-evolution for these two cell types. Moreover, due to the low probability of two adjacent cells simultaneously acquiring genetic changes, this would also suggest that breast cancer epithelial and stromal cells may be derived from a common stem cell and then undergo a divergent genetic selection process.

In order to determine if in the same population of tumor epithelial, myoepithelial, and stromal cells in which we detected dramatic gene expression changes there are also underlying genetic alterations, we analyzed the genotype of these cell types using different technologies in 2 DCIS and 12 invasive breast carcinomas. All but one of the tumor epithelial cells had numerous LOH involving almost all chromosome arms. The most frequent genetic changes we identified in the tumor epithelial cells (1q, 8q, 17q, and 20q gain, and 6q, 8p, 10q, 12q, and 17p LOH), both in DCIS and invasive tumors, were in agreement with that of prior studies (Nishizaki et al., 1997; Waldman et al., 2000). The one tumor DNA sample (IDC10) that appeared to be devoid of significant LOH was obtained from a low-grade estrogen and progesterone receptor-positive, HER2-negative breast tumor. The lack of gross chromosomal changes in this tumor is unlikely to be due to technical issues, but potentially reflects a special pathway of breast tumorigenesis. Correlating with this, an independent study using BAC array CGH analysis of a large set of breast tumors also found that a subset of breast tumors (9/146) have minimal (<1.5% of the genome) chromosomal changes (Dr. J. Gray, Lawrence Berkeley National Laboratory, personal communication). Using three different methods (aCGH, SNP arrays, and direct sequencing of specific SNPs), we detected genetic changes only in cancer epithelial cells. However, in a malignant phyllodes tumor that is thought to be composed of malignant stroma and epithelium, we detected LOH in both components. These results suggest that using the technologies we applied, genetic changes can be detected both in epithelial and stromal cells, but only if there is a mono- or oligoclonal proliferation of neoplastic epithelial or stromal cells. Our inability to find conclusive genetic alterations in stromal cells from invasive ductal breast carcinomas is seemingly in disagreement with the findings of several of the above referenced studies. However, we believe that the reason for the different results could be due to the use of different technologies and approaches. All the studies that described LOH in cancer stroma analyzed a few polymorphic markers and a fairly small population of stromal cells isolated by microdissection from the same area adjacent to tumor epithelial cells, while we analyzed all the stromal cells from the tumor and used comprehensive genome-wide SNP arrays. Thus, if the stromal cells are highly heterogeneous with respect to genetic alterations, these changes can be detected only if relatively few cells from the same area of the tumor are analyzed. However, in our view this argues against the hypothesis that the genetic changes in the stroma are selected for and thus play a major role in tumorigenesis.

In summary, this study provides a comprehensive molecular characterization of each cell type composing normal breast tissue and in situ and invasive breast carcinomas. The genes described here should therefore provide a valuable resource for future basic and clinical studies addressing the role of epithelial-stromal cell interactions in breast and other cancer types. The availability of specific chemokine receptor inhibitors and preclinical studies demonstrating dramatic tumor and metastasis suppressive effects using CXCR4 inhibitors in brain and breast tumors (Rubin et al., 2003; Tamamura et al., 2003) provide a proof of principle that therapeutic targeting of chemokines is a promising new opportunity for the treatment of breast carcinomas.

Experimental procedures

Cell lines and tissue specimens

Breast cancer cell lines were obtained from American Type Culture Collection (Manassas, VA) or were generously provided by Drs. Steve Ethier (University of Michigan) and Arthur Pardee (Dana-Farber Cancer Institute). Cells were grown in media recommended by the provider. Tumor specimens were obtained from Brigham and Women's and Massachusetts General Hospitals (Boston, MA), Duke University (Durham, NC), University Hospital Zagreb (Zagreb, Croatia), and the National Disease Research Interchange, snap frozen on dry ice, and stored at −80°C until use or were processed for purification as described below. All human tissue was collected using protocols approved by the Institutional Review Boards. We purified all the cell types from 2 different normal reduction mammoplasty tissues, 2 different DCIS, and 13 different invasive ductal carcinomas. Due to technical difficulties (insufficient number of cells), we were not able to generate SAGE libraries from each cell type of each tissue used for purification. In addition, selected cell types were isolated from a few additional normal and DCIS samples. The detailed protocol used for the purification of all cell types is included in the Supplemental Data. The estimated number of cells obtained from each fraction varied from 10,000 to 100, 000.

Generation and analysis of SAGE libraries, mRNA in situ hybridization, and immunohistochemistry

All SAGE libraries were generated using a modified micro-SAGE protocol and the I-SAGE (libraries prepared in 2002) or long I-SAGE (I-epi-7, I-epi-8, I-epi-9, I-leu-7, I-str-7, I-myofib-7, I-myofib-8, I-myofib-9, D-str-6, FA, PHY) kits from Invitrogen. The samples were collected and SAGE libraries were generated during 2002–2004, and the long-SAGE kit became available only in 2003. SAGE libraries were sequenced by Agencourt (Beverly, MA) as part of the NCI-CGAP SAGE project, and all data will be deposited to the SAGEGenie website (http://cgap.nci.nih.gov/SAGE). Approximately 50,000 tags (average tag number 56,647 ± 4,383) were obtained from each library, and the preliminary analysis of the SAGE data was performed essentially as described (Porter et al., 2001). Briefly, genes significantly ($p \leq 0.002$) differentially expressed between normal and cancerous cells were identified by performing pair-wise comparisons using the SAGE2000 software and Monte Carlo analysis. Significance calculation among groups of SAGE libraries and clustering analyses were performed using a new Poisson model-based K-means algorithm (PK algorithm, Cai et al., 2004). A detailed descrip-

tion of the methodology used for the analysis and clustering of the SAGE data is provided in the Supplemental Data. Probes for the selected genes to be used for mRNA in situ hybridization were generated by PCR amplification of a 300–500 bp region of the 3′UTR and subcloning the fragments into pZERO1.0 (Invitrogen). The identity of the subcloned PCR products was confirmed by sequencing, and the resulting plasmids were used for the generation of digitonin-labeled riboprobes essentially as described (Porter et al., 2003a). mRNA in situ hybridizations and immunohistochemistry were performed as described or as recommended by the antibody supplier (Porter et al., 2003a). Mouse monoclonal antibodies for IL1β and CCL3 were purchased from R&D, while anti-CD45 and anti-Ki67 mouse monoclonal antibodies were obtained from DAKO.

Array comparative genomic hybridization

cDNA array comparative genomic hybridization using Agilent (Palo Alto, California) arrays were performed by the Belfer Genome Center at the Dana-Farber Cancer Institute. Genomic DNA was digested with DpnII and random prime labeled according to standard protocols with slight modifications (Pollack et al., 1999). (For a detailed protocol, see http://genomic.dfci.harvard.edu/array_cgh.htm.) Labeled DNAs were hybridized to human cDNA microarrays containing 12,814 unique cDNA clones (Agilent Technologies, Human 1 clone set). Among these clones, approximately 9,420 unique map positions were found for 12,020 unique GenBank sequences. The median interval between cDNAs is 100.1 kilobase, 92.8% of intervals are less than 1 megabase, and 98.6% are less than 3 megabases. The density of coverage is closely correlated with gene density. Following extensive QA analysis, fluorescence ratios of scanned images of the arrays were calculated and the raw array CGH profiles were processed to identify statistically significant transitions in copy number using a segmentation algorithm that employs permutation to determine the significance of change points in the raw data. By mode centering this segmented data set, we defined gains and losses as Log2 signal ratio of greater than or equal to +0.13 or −0.13, respectively, and amplification and deletion as a ratio greater than 0.52 or less than −0.58, respectively (e.g., 97% or 3% quantiles). Statistical analysis of the aCGH data will be described in detail elsewhere (Brennan et al., 2004). Segmentation of aCGH profiles was performed by changepoint identification algorithm provided by Adam Olshen and E.S. Venkatraman (Lucito et al., 2003).

Single nucleotide polymorphism array analysis

SNP array hybridizations were performed by the Dana-Farber Microarray Core using Affymetrix 11K XbaI SNP arrays and protocols recommended by Affymetrix (Santa Clara, CA). These arrays contain probes for both alleles at 11,565 SNP loci, with mean and median intermarker distances of 209 kb and 104 kb, respectively, with probe density closely correlating with gene density (Matsuzaki et al., 2004). Arrays were scanned using a confocal laser scanner (Agilent, Palo Alto, California), and Affymetrix genotyping software (Affymetrix GeneChip 5.0) was used to make allele calls for all loci. These data were then analyzed using dChipSNP (Lieberfarb et al., 2003; Lin et al., 2004). Loci that were heterozygous (AB call) in normal epithelium or leukocytes but homozygous (AA or BB call) in the test tissue were identified as potentially having undergone LOH. Some of these potential LOH events reflect genotyping error, but serial events among neighboring loci along a chromosome likely reflect true regional LOH. Regions of statistically likely LOH were delineated according to a Hidden Markov Model analysis; the detailed method will be described elsewhere. Two samples (LCIS and FA) had no normal reference counterparts. For these, regions of LOH were inferred when a stretch of consecutive homozygous loci exceeded what would be expected by chance alone. Again, a Hidden Markov Model analysis was used, assigning a marginal probability of heterozygosity of 0.37 to correspond to the actual rate found in these SNPs (Matsuzaki et al., 2004) and a transition probability between consecutive SNPs proportional to the genetic distance between them (Lander and Green, 1987). A detailed description of the method will be presented in a forthcoming study (M. Lin, R.B., X. Zhao, M. Meyerson, C. Li, and W.R.S., unpublished data). Samples were clustered hierarchically as previously described (Lin et al., 2004), based upon LOH calls in all statistically likely regions of LOH in all chromosomes.

Ligand binding, cell growth, migration, and invasion assays

We generated N-terminal or C-terminal alkaline phosphatase (AP) CXCL14 fusion proteins using the AP-TAG-5 expression vector (GenHunter, Nashville, TN). We transfected mammalian cells with Fugene6 (Roche, Indianapolis, IN), Lipofectamine, or Lipofectamine 2000 (LifeTechnologies, Rockville, MD) reagents. We performed in vivo and in vitro ligand binding assays on primary tissues and cell lines using AP-CXCL14 essentially as described (Flanagan and Leder, 1990; Porter et al., 2003b). Briefly, we fixed frozen sections of various human specimens, incubated with either AP-CXCL14 fusion protein or AP control conditioned medium, rinsed, and then incubated with AP substrate forming a blue/purple precipitate. For in vitro assays we incubated cells in suspension with conditioned media containing either AP alone or AP-CXCL14 fusion protein, rinsed, and then assayed for bound AP activity. To determine the effect of CXCL14 on cell growth, we plated MDA-MB-231 and MCF10A cells (4000 cells/well in a 24-well plate) and grew them in conditioned media containing AP or AP-CXCL14. Conditioned media were generated by transfecting 293 cells with pAP-tag5 or pAP-CXCL14 plasmids and growing them in McCoy's medium supplemented with 10% FBS (to be used for MDA-MB-231 cells) or in MCF10A media (American Type Culture Collection, Manassas, VA, to be used in MCF10A cells). Cells were counted (3 wells/time point) on days 1, 2, 4, 6, and 8 after plating. We used 10 nM CXCL12 in MDA-MB-231 cells as positive control. The experiment was repeated three times. In order to determine if CXCL14 binding to breast cancer cells has an effect on cell migration and invasion, we tested the ability of conditioned medium containing AP-CXCL14 or pCDNA3.1 expressing HA-tagged CXCL14 to induce the migration and invasion of MDA-MB-231 cells using BIOCOAT Matrigel invasion chambers essentially as described (Muller et al., 2001). For invasion assays, we plated 2.5×10^4 cells/well and assayed 24 hr later, while for migration assays we used 1.25×10^4 cells/well and determined cell numbers 12 hr later. Conditioned medium of cells transfected with pAP-Tag5 or pCDNA 3.1 empty vectors were used as negative control.

Acknowledgments

We thank Gabriela Lodeiro and Ruth Gomes for their help with the acquisition of tumor samples and Drs. Barrett Rollins and Myles Brown for critically reading the manuscript. This work is funded in part by NIH (NCI-CGAP SAGE and SPORE grant P50 CA89393-01 to K.P. and NRSA Fellowship 5F32CA94788-02 to D.P.) and DOD (grant DAMD 17 01 1 0221 to M.A.). W.R.S. and K.P. receive research support from and are consultants to the Novartis Institutes for Biomedical Research. K.P. is a consultant to GenPath Pharmaceuticals, Inc.

Received: November 12, 2003
Revised: March 16, 2004
Accepted: May 18, 2004
Published: July 19, 2004

References

Alpaugh, M.L., Lee, M.C., Nguyen, M., Deato, M., Dishakjian, L., and Barsky, S.H. (2000). Myoepithelial-specific CD44 shedding contributes to the anti-invasive and antiangiogenic phenotype of myoepithelial cells. Exp. Cell Res. *261*, 150–158.

Amiel, A., Kaufman, Z., Goldstein, E., Bruchim, R.B., Kidron, D., Gaber, E., and Fejgin, M.D. (2003). Application of comparative genomic hybridization in search for genetic aberrations in fibroadenomas of the breast. Cancer Genet. Cytogenet. *142*, 145–148.

Barbero, S., Bajetto, A., Bonavia, R., Porcile, C., Piccioli, P., Pirani, P., Ravetti, J.L., Zona, G., Spaziante, R., Florio, T., and Schettini, G. (2002). Expression of the chemokine receptor CXCR4 and its ligand stromal cell-derived factor 1 in human brain tumors and their involvement in glial proliferation in vitro. Ann. N Y Acad. Sci. *973*, 60–69.

Barsky, S.H. (2003). Myoepithelial mRNA expression profiling reveals a common tumor-suppressor phenotype. Exp. Mol. Pathol. *74*, 113–122.

Bhowmick, N.A., Chytil, A., Plieth, D., Gorska, A.E., Dumont, N., Shappell, S., Washington, M.K., Neilson, E.G., and Moses, H.L. (2004). TGF-beta signaling in fibroblasts modulates the oncogenic potential of adjacent epithelia. Science *303*, 848–851.

Bissell, M.J., and Radisky, D. (2001). Putting tumours in context. Nat. Rev. Cancer *1*, 46–54.

Bissell, M.J., Radisky, D.C., Rizki, A., Weaver, V.M., and Petersen, O.W. (2002). The organizing principle: microenvironmental influences in the normal and malignant breast. Differentiation *70*, 537–546.

Bocker, W., Moll, R., Poremba, C., Holland, R., Van Diest, P.J., Dervan, P., Burger, H., Wai, D., Ina Diallo, R., Brandt, B., et al. (2002). Common adult stem cells in the human breast give rise to glandular and myoepithelial cell lineages: a new cell biological concept. Lab. Invest. *82*, 737–746.

Brennan, C., Zhang, Y., Leo, C., Feng, B., Cauwels, C., Aguirre, A.J., Kim, M., Protopopov, A., and Chin, L. High-resolution global profiling of genomic alterations with long oligonucleotide microarray. Cancer Res., in press.

Cai, L., Huang, H., Blackshaw, S., Liu, J.S., Cepko, C., and Wong, W.H. (2004). Clustering analysis of SAGE data using a Poisson approach. Genome Biol., in press.

Chauhan, H., Abraham, A., Phillips, J.R., Pringle, J.H., Walker, R.A., and Jones, J.L. (2003). There is more than one kind of myofibroblast: analysis of CD34 expression in benign, in situ, and invasive breast lesions. J. Clin. Pathol. *56*, 271–276.

Chen, Y., Stamatoyannopoulos, G., and Song, C.Z. (2003). Down-regulation of CXCR4 by inducible small interfering RNA inhibits breast cancer cell invasion in vitro. Cancer Res. *63*, 4801–4804.

Coussens, L.M., and Werb, Z. (2002). Inflammation and cancer. Nature *420*, 860–867.

DeCosse, J.J., Gossens, C.L., Kuzma, J.F., and Unsworth, B.R. (1973). Breast cancer: induction of differentiation by embryonic tissue. Science *181*, 1057–1058.

DeCosse, J.J., Gossens, C., Kuzma, J.F., and Unsworth, B.R. (1975). Embryonic inductive tissues that cause histologic differentiation of murine mammary carcinoma in vitro. J. Natl. Cancer Inst. *54*, 913–922.

Deng, G., Lu, Y., Zlotnikov, G., Thor, A.D., and Smith, H.S. (1996). Loss of heterozygosity in normal tissue adjacent to breast carcinomas. Science *274*, 2057–2059.

Deugnier, M.A., Teuliere, J., Faraldo, M.M., Thiery, J.P., and Glukhova, M.A. (2002). The importance of being a myoepithelial cell. Breast Cancer Res. *4*, 224–230.

De Wever, O., and Mareel, M. (2003). Role of tissue stroma in cancer cell invasion. J. Pathol. *200*, 429–447.

Dontu, G., Abdallah, W.M., Foley, J.M., Jackson, K.W., Clarke, M.F., Kawamura, M.J., and Wicha, M.S. (2003). In vitro propagation and transcriptional profiling of human mammary stem/progenitor cells. Genes Dev. *17*, 1253–1270.

Elenbaas, B., and Weinberg, R.A. (2001). Heterotypic signaling between epithelial tumor cells and fibroblasts in carcinoma formation. Exp. Cell Res. *264*, 169–184.

Flanagan, J.G., and Leder, P. (1990). The kit ligand: a cell surface molecule altered in steel mutant fibroblasts. Cell *63*, 185–194.

Forsti, A., Louhelainen, J., Soderberg, M., Wijkstrom, H., and Hemminki, K. (2001). Loss of heterozygosity in tumour-adjacent normal tissue of breast and bladder cancer. Eur. J. Cancer *37*, 1372–1380.

Frederick, M.J., Henderson, Y., Xu, X., Deavers, M.T., Sahin, A.A., Wu, H., Lewis, D.E., El-Naggar, A.K., and Clayman, G.L. (2000). In vivo expression of the novel CXC chemokine BRAK in normal and cancerous human tissue. Am. J. Pathol. *156*, 1937–1950.

Gabbiani, G. (1999). Some historical and philosophical reflections on the myofibroblast concept. Curr. Top. Pathol. *93*, 1–5.

Gerard, C., and Rollins, B.J. (2001). Chemokines and disease. Nat. Immunol. *2*, 108–115.

Gomm, J.J., Browne, P.J., Coope, R.C., Bansal, G.S., Yiangou, C., Johnston, C.L., Mason, R., and Coombes, R.C. (1997a). A paracrine role for myoepithelial cell-derived FGF2 in the normal human breast. Exp. Cell Res. *234*, 165–173.

Gomm, J.J., Coope, R.C., Browne, P.J., and Coombes, R.C. (1997b). Separated human breast epithelial and myoepithelial cells have different growth factor requirements in vitro but can reconstitute normal breast lobuloalveolar structure. J. Cell. Physiol. *171*, 11–19.

Gudjonsson, T., Ronnov-Jessen, L., Villadsen, R., Rank, F., Bissell, M.J., and Petersen, O.W. (2002). Normal and tumor-derived myoepithelial cells differ in their ability to interact with luminal breast epithelial cells for polarity and basement membrane deposition. J. Cell Sci. *115*, 39–50.

Hall, J.M., and Korach, K.S. (2003). Stromal cell-derived factor *1*, a novel target of estrogen receptor action, mediates the mitogenic effects of estradiol in ovarian and breast cancer cells. Mol. Endocrinol. *17*, 792–803.

Hromas, R., Broxmeyer, H.E., Kim, C., Nakshatri, H., Christopherson, K., 2nd, Azam, M., and Hou, Y.H. (1999). Cloning of BRAK, a novel divergent CXC chemokine preferentially expressed in normal versus malignant cells. Biochem. Biophys. Res. Commun. *255*, 703–706.

Hunter, K.W. (2004). Host genetics and tumour metastasis. Br. J. Cancer *90*, 752–755.

Ishii, G., Sangai, T., Oda, T., Aoyagi, Y., Hasebe, T., Kanomata, N., Endoh, Y., Okumura, C., Okuhara, Y., Magae, J., et al. (2003). Bone-marrow-derived myofibroblasts contribute to the cancer-induced stromal reaction. Biochem. Biophys. Res. Commun. *309*, 232–240.

Kang, Y., Siegel, P.M., Shu, W., Drobnjak, M., Kakonen, S.M., Cordon-Cardo, C., Guise, T.A., and Massague, J. (2003). A multigenic program mediating breast cancer metastasis to bone. Cancer Cell *3*, 537–549.

Kenny, P.A., and Bissell, M.J. (2003). Tumor reversion: correction of malignant behavior by microenvironmental cues. Int. J. Cancer *107*, 688–695.

Kurose, K., Hoshaw-Woodard, S., Adeyinka, A., Lemeshow, S., Watson, P.H., and Eng, C. (2001). Genetic model of multi-step breast carcinogenesis involving the epithelium and stroma: clues to tumour-microenvironment interactions. Hum. Mol. Genet. *10*, 1907–1913.

Kurose, K., Gilley, K., Matsumoto, S., Watson, P.H., Zhou, X.P., and Eng, C. (2002). Frequent somatic mutations in PTEN and TP53 are mutually exclusive in the stroma of breast carcinomas. Nat. Genet. *32*, 355–357.

Kurth, I., Willimann, K., Schaerli, P., Hunziker, T., Clark-Lewis, I., and Moser, B. (2001). Monocyte selectivity and tissue localization suggests a role for breast and kidney-expressed chemokine (BRAK) in macrophage development. J. Exp. Med. *194*, 855–861.

Lakhani, S.R., Jacquemier, J., Sloane, J.P., Gusterson, B.A., Anderson, T.J., van de Vijver, M.J., Farid, L.M., Venter, D., Antoniou, A., Storfer-Isser, A., et al. (1998). Multifactorial analysis of differences between sporadic breast cancers and cancers involving BRCA1 and BRCA2 mutations. J. Natl. Cancer Inst. *90*, 1138–1145.

Lakhani, S.R., Chaggar, R., Davies, S., Jones, C., Collins, N., Odel, C., Stratton, M.R., and O'Hare, M.J. (1999). Genetic alterations in 'normal' luminal and myoepithelial cells of the breast. J. Pathol. *189*, 496–503.

Lander, E.S., and Green, P. (1987). Construction of multilocus genetic linkage maps in humans. Proc. Natl. Acad. Sci. USA *84*, 2363–2367.

Lieberfarb, M.E., Lin, M., Lechpammer, M., Li, C., Tanenbaum, D.M., Febbo, P.G., Wright, R.L., Shim, J., Kantoff, P.W., Loda, M., et al. (2003). Genome-wide loss of heterozygosity analysis from laser capture microdissected prostate cancer using single nucleotide polymorphic allele (SNP) arrays and a novel bioinformatics platform dChipSNP. Cancer Res. *63*, 4781–4785.

Lin, M., Wei, L.J., Sellers, W.R., Lieberfarb, M., Wong, W.H., and Li, C. (2004). dChipSNP: significance curve and clustering of SNP-array-based loss-of-heterozygosity data. Bioinformatics *20*, 1233–1240.

Lucito, R., Healy, J., Alexander, J., Reiner, A., Esposito, D., Chi, M., Rodgers, L., Brady, A., Sebat, J., Troge, J., et al. (2003). Representational oligonucleotide microarray analysis: a high-resolution method to detect genome copy number variation. Genome Res. *13*, 2291–2305.

Maffini, M.V., Soto, A.M., Calabro, J.M., Ucci, A.A., and Sonnenschein, C. (2004). The stroma as a crucial target in rat mammary gland carcinogenesis. J. Cell Sci. *117*, 1495–1502.

Man, Y.G., Tai, L., Barner, R., Vang, R., Saenger, J.S., Shekitka, K.M., Brat-

thauer, G.L., Wheeler, D.T., Liang, C.Y., Vinh, T.N., and Strauss, B.L. (2003). Cell clusters overlying focally disrupted mammary myoepithelial cell layers and adjacent cells within the same duct display different immunohistochemical and genetic features: implications for tumor progression and invasion. Breast Cancer Res. *5*, R231–R241.

Matsuzaki, H., Loi, H., Dong, S., Tsai, Y.Y., Fang, J., Law, J., Di, X., Liu, W.M., Yang, G., Liu, G., et al. (2004). Parallel genotyping of over 10,000 SNPs using a one-primer assay on a high-density oligonucleotide array. Genome Res. *14*, 414–425.

Moinfar, F., Man, Y.G., Arnould, L., Bratthauer, G.L., Ratschek, M., and Tavassoli, F.A. (2000). Concurrent and independent genetic alterations in the stromal and epithelial cells of mammary carcinoma: implications for tumorigenesis. Cancer Res. *60*, 2562–2566.

Muller, A., Homey, B., Soto, H., Ge, N., Catron, D., Buchanan, M.E., McClanahan, T., Murphy, E., Yuan, W., Wagner, S.N., et al. (2001). Involvement of chemokine receptors in breast cancer metastasis. Nature *410*, 50–56.

Murrell, T.G. (1995). The potential for oxytocin (OT) to prevent breast cancer: a hypothesis. Breast Cancer Res. Treat. *35*, 225–229.

Nguyen, M., Lee, M.C., Wang, J.L., Tomlinson, J.S., Shao, Z.M., Alpaugh, M.L., and Barsky, S.H. (2000). The human myoepithelial cell displays a multifaceted anti-angiogenic phenotype. Oncogene *19*, 3449–3459.

Nishizaki, T., DeVries, S., Chew, K., Goodson, W.H., 3rd, Ljung, B.M., Thor, A., and Waldman, F.M. (1997). Genetic alterations in primary breast cancers and their metastases: direct comparison using modified comparative genomic hybridization. Genes Chromosomes Cancer *19*, 267–272.

Oh, S.J., Kurz, H., Christ, B., and Wilting, J. (1998). Platelet-derived growth factor-B induces transformation of fibrocytes into spindle-shaped myofibroblasts in vivo. Histochem. Cell Biol. *109*, 349–357.

Page, M.J., Amess, B., Townsend, R.R., Parekh, R., Herath, A., Brusten, L., Zvelebil, M.J., Stein, R.C., Waterfield, M.D., Davies, S.C., and O'Hare, M.J. (1999). Proteomic definition of normal human luminal and myoepithelial breast cells purified from reduction mammoplasties. Proc. Natl. Acad. Sci. USA *96*, 12589–12594.

Pollack, J.R., Perou, C.M., Alizadeh, A.A., Eisen, M.B., Pergamenschikov, A., Williams, C.F., Jeffrey, S.S., Botstein, D., and Brown, P.O. (1999). Genome-wide analysis of DNA copy-number changes using cDNA microarrays. Nat. Genet. *23*, 41–46.

Porter, D.A., Krop, I.E., Nasser, S., Sgroi, D., Kaelin, C.M., Marks, J.R., Riggins, G., and Polyak, K. (2001). A SAGE (serial analysis of gene expression) view of breast tumor progression. Cancer Res. *61*, 5697–5702.

Porter, D., Lahti-Domenici, J., Keshaviah, A., Bae, Y.K., Argani, P., Marks, J., Richardson, A., Cooper, A., Strausberg, R., Riggins, G.J., et al. (2003a). Molecular markers in ductal carcinoma in situ of the breast. Mol. Cancer Res. *1*, 362–375.

Porter, D., Weremowicz, S., Chin, K., Seth, P., Keshaviah, A., Lahti-Domenici, J., Bae, Y.K., Monitto, C.L., Merlos-Suarez, A., Chan, J., et al. (2003b). A neural survival factor is a candidate oncogene in breast cancer. Proc. Natl. Acad. Sci. USA *100*, 10931–10936.

Radisky, D., Hagios, C., and Bissell, M.J. (2001). Tumors are unique organs defined by abnormal signaling and context. Semin. Cancer Biol. *11*, 87–95.

Rajagopalan, H., Nowak, M.A., Vogelstein, B., and Lengauer, C. (2003). The significance of unstable chromosomes in colorectal cancer. Nat. Rev. Cancer *3*, 695–701.

Ramaswamy, S., Ross, K.N., Lander, E.S., and Golub, T.R. (2003). A molecular signature of metastasis in primary solid tumors. Nat. Genet. *33*, 49–54.

Rohan, R.M., Fernandez, A., Udagawa, T., Yuan, J., and D'Amato, R.J. (2000). Genetic heterogeneity of angiogenesis in mice. FASEB J. *14*, 871–876.

Ronnov-Jessen, L., Petersen, O.W., Koteliansky, V.E., and Bissell, M.J. (1995). The origin of the myofibroblasts in breast cancer. Recapitulation of tumor environment in culture unravels diversity and implicates converted fibroblasts and recruited smooth muscle cells. J. Clin. Invest. *95*, 859–873.

Ronnov-Jessen, L., Petersen, O.W., and Bissell, M.J. (1996). Cellular changes involved in conversion of normal to malignant breast: importance of the stromal reaction. Physiol. Rev. *76*, 69–125.

Rossi, D., and Zlotnik, A. (2000). The biology of chemokines and their receptors. Annu. Rev. Immunol. *18*, 217–242.

Rubin, J.B., Kung, A.L., Klein, R.S., Chan, J.A., Sun, Y., Schmidt, K., Kieran, M.W., Luster, A.D., and Segal, R.A. (2003). A small-molecule antagonist of CXCR4 inhibits intracranial growth of primary brain tumors. Proc. Natl. Acad. Sci. USA *100*, 13513–13518.

Sawyer, E.J., Hanby, A.M., Ellis, P., Lakhani, S.R., Ellis, I.O., Boyle, S., and Tomlinson, I.P. (2000). Molecular analysis of phyllodes tumors reveals distinct changes in the epithelial and stromal components. Am. J. Pathol. *156*, 1093–1098.

Sawyer, E.J., Hanby, A.M., Rowan, A.J., Gillett, C.E., Thomas, R.E., Poulsom, R., Lakhani, S.R., Ellis, I.O., Ellis, P., and Tomlinson, I.P. (2002). The Wnt pathway, epithelial-stromal interactions, and malignant progression in phyllodes tumours. J. Pathol. *196*, 437–444.

Schena, M., Shalon, D., Davis, R.W., and Brown, P.O. (1995). Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science *270*, 467–470.

Schurch, W. (1999). The myofibroblast in neoplasia. Curr. Top. Pathol. *93*, 135–148.

Scotton, C.J., Wilson, J.L., Scott, K., Stamp, G., Wilbanks, G.D., Fricker, S., Bridger, G., and Balkwill, F.R. (2002). Multiple actions of the chemokine CXCL12 on epithelial tumor cells in human ovarian cancer. Cancer Res. *62*, 5930–5938.

Shao, Z.M., Nguyen, M., Alpaugh, M.L., O'Connell, J.T., and Barsky, S.H. (1998). The human myoepithelial cell exerts antiproliferative effects on breast carcinoma cells characterized by p21WAF1/CIP1 induction, G2/M arrest, and apoptosis. Exp. Cell Res. *241*, 394–403.

Shekhar, M.P., Pauley, R., and Heppner, G. (2003). Host microenvironment in breast cancer development: extracellular matrix-stromal cell contribution to neoplastic phenotype of epithelial cells in the breast. Breast Cancer Res. *5*, 130–135.

Sleeman, M.A., Fraser, J.K., Murison, J.G., Kelly, S.L., Prestidge, R.L., Palmer, D.J., Watson, J.D., and Kumble, K.D. (2000). B cell- and monocyte-activating chemokine (BMAC), a novel non-ELR alpha-chemokine. Int. Immunol. *12*, 677–689.

Sorlie, T., Perou, C.M., Tibshirani, R., Aas, T., Geisler, S., Johnsen, H., Hastie, T., Eisen, M.B., van de Rijn, M., Jeffrey, S.S., et al. (2001). Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc. Natl. Acad. Sci. USA *98*, 10869–10874.

St Croix, B., Rago, C., Velculescu, V., Traverso, G., Romans, K.E., Montgomery, E., Lal, A., Riggins, G.J., Lengauer, C., Vogelstein, B., and Kinzler, K.W. (2000). Genes expressed in human tumor endothelium. Science *289*, 1197–1202.

Sternlicht, M.D., and Barsky, S.H. (1997). The myoepithelial defense: a host defense against cancer. Med. Hypotheses *48*, 37–46.

Sternlicht, M.D., Kedeshian, P., Shao, Z.M., Safarians, S., and Barsky, S.H. (1997). The human myoepithelial cell is a natural tumor suppressor. Clin. Cancer Res. *3*, 1949–1958.

Tamamura, H., Hori, A., Kanzaki, N., Hiramatsu, K., Mizumoto, M., Nakashima, H., Yamamoto, N., Otaka, A., and Fujii, N. (2003). T140 analogs as CXCR4 antagonists identified as anti-metastatic agents in the treatment of breast cancer. FEBS Lett. *550*, 79–83.

Tanaka, K., Sano, K., Yuba, K., Katsumura, K., Nakano, T., Kobayashi, M., Ikeda, T., and Abe, M. (2003). Inhibition of induction of myofibroblasts by interferon gamma in a human fibroblast cell line. Int. Immunopharmacol. *3*, 1273–1280.

Tlsty, T.D. (2001). Stromal cells can contribute oncogenic signals. Semin. Cancer Biol. *11*, 97–104.

Tlsty, T.D., and Hein, P.W. (2001). Know thy neighbor: stromal cells can contribute oncogenic signals. Curr. Opin. Genet. Dev. *11*, 54–59.

van 't Veer, L.J., Dai, H., van de Vijver, M.J., He, Y.D., Hart, A.A., Mao, M.,

Peterse, H.L., van der Kooy, K., Marton, M.J., Witteveen, A.T., et al. (2002). Gene expression profiling predicts clinical outcome of breast cancer. Nature *415*, 530–536.

van de Vijver, M.J., He, Y.D., van 't Veer, L.J., Dai, H., Hart, A.A., Voskuil, D.W., Schreiber, G.J., Peterse, J.L., Roberts, C., Marton, M.J., et al. (2002). A gene-expression signature as a predictor of survival in breast cancer. N. Engl. J. Med. *347*, 1999–2009.

Velculescu, V.E., Zhang, L., Vogelstein, B., and Kinzler, K.W. (1995). Serial analysis of gene expression. Science *270*, 484–487.

Waldman, F.M., DeVries, S., Chew, K.L., Moore, D.H., 2nd, Kerlikowske, K., and Ljung, B.M. (2000). Chromosomal alterations in ductal carcinomas in situ and their in situ recurrences. J. Natl. Cancer Inst. *92*, 313–320.

Walter-Yohrling, J., Pratt, B.M., Ledbetter, S., and Teicher, B.A. (2003). Myofibroblasts enable invasion of endothelial cells into three-dimensional tumor cell clusters: a novel in vitro tumor model. Cancer Chemother. Pharmacol. *52*, 263–269.

Wernert, N., Locherbach, C., Wellmann, A., Behrens, P., and Hugel, A. (2001). Presence of genetic alterations in microdissected stroma of human colon and breast cancers. Anticancer Res. *21*, 2259–2264.

Wiseman, B.S., and Werb, Z. (2002). Stromal effects on mammary gland development and breast cancer. Science *296*, 1046–1049.

Xiao, G., Liu, Y.E., Gentz, R., Sang, Q.A., Ni, J., Goldberg, I.D., and Shi, Y.E. (1999). Suppression of breast cancer growth and metastasis by a serpin myoepithelium-derived serine proteinase inhibitor expressed in the mammary myoepithelial cells. Proc. Natl. Acad. Sci. USA *96*, 3700–3705.

Zhang, R.R., Man, Y.G., Vang, R., Saenger, J.S., Barner, R., Wheeler, D.T., Liang, C.Y., Vinh, T.N., and Bratthauer, G.L. (2003). A subset of morphologically distinct mammary myoepithelial cells lacks corresponding immunophenotypic markers. Breast Cancer Res. *5*, R151–R156.

Zhu, G., Reynolds, L., Crnogorac-Jurcevic, T., Gillett, C.E., Dublin, E.A., Marshall, J.F., Barnes, D., D'Arrigo, C., Van Trappen, P.O., Lemoine, N.R., and Hart, I.R. (2003). Combination of microdissection and microarray analysis to identify gene expression changes between differentially located tumour cells in breast cancer. Oncogene *22*, 3742–3748.

Oncogene (2002) 21, 836–843
© 2002 Nature Publishing Group All rights reserved 0950–9232/02 $25.00
www.nature.com/onc

SHORT REPORTS

Novel estrogen and tamoxifen induced genes identified by SAGE (Serial Analysis of Gene Expression)

Pankaj Seth[1,2], Ian Krop[1,2], Dale Porter[1,2] and Kornelia Polyak*[,1,2]

[1]*Department of Adult Oncology, Dana-Farber Cancer Institute, Boston, Massachusetts, MA 02115, USA;* [2]*Department of Medicine, Harvard Medical School, Boston, Massachusetts, MA 02115, USA*

The breast cancer promoting effects of estrogen and the chemopreventive effects of tamoxifen are thought to be mediated by the estrogen receptor, a ligand-dependent transcription factor. Therefore, comprehensive analysis of gene expression profiles following estrogen or tamoxifen treatment may help us better understand the role estrogen plays in tumorigenesis. We utilized SAGE (Serial Analysis of Gene Expression) technology to identify genes regulated by estrogen and tamoxifen in the ZR75-1 estrogen dependent breast cancer cell line. In this manner we have identified several genes that were regulated by estrogen or tamoxifen. Here we report the identification and initial characterization of EIT-6 (Estrogen Induced Tag-6), a novel nuclear protein and a new member of the evolutionarily conserved SM-20 family of growth regulatory immediate-early genes. EIT-6 appears to be a direct transcriptional target of the estrogen receptor and constitutive expression of EIT-6 promotes colony growth in human breast cancer cells. These data indicate that EIT-6 may play a role in estrogen induced cell growth.
Oncogene (2002) **21,** 836–843. DOI: 10.1038/sj/onc/1205113

Keywords: estrogen; tamoxifen; SAGE (Serial Analysis of Gene Expression); breast cancer

Breast cancer is a leading cause of cancer death in women of the Western world (Greenlee *et al.*, 2000). Despite advances in early detection and treatment, breast cancer mortality rates have not decreased significantly over the past few decades. Thus, there is a continued and increasing need for the identification of risk factors of breast cancer and molecular targets of chemoprevention in order to decrease the incidence of this disease. Estrogen plays a key role in the development of the normal mammary gland and in the initiation and progression of breast carcinomas (Nandi *et al.*, 1995). Clinical trials using anti-estrogens (tamoxifen) proved the importance of estrogens in breast tumor development, and identified tamoxifen as a breast cancer preventive agent (Fisher *et al.*, 1998, 1999). However, there is little known about the mechanisms that account for the tumorigenic effects of estrogen and the cancer preventive effects of tamoxifen.

The action of estrogen is mediated by its receptors (estrogen receptors-ERα and β), members of the nuclear hormone receptor family and ligand-dependent transcription factors (Katzenellenbogen, 1996; Mangelsdorf *et al.*, 1995). Estrogen binding stimulates the trans-activating function of ER through its ability to facilitate the recruitment of various receptor binding co-activator proteins (Freedman, 1999). These receptor-co-activator complexes then affect transcription initiation at promoters regulated by estrogen. Anti-estrogens not only preclude co-activator binding, but can facilitate the recruitment of co-repressors and lead to active repression of the basal expression of certain genes. Many of the co-activators and co-repressors are cell type and differentiation stage specific, therefore their interaction with ER may explain the diverse, sometimes opposing effects of estrogen and tamoxifen in different cell types. Several studies have been performed to identify genes whose expression is modulated by estrogen or tamoxifen treatment (Charpentier *et al.*, 2000; Inadera *et al.*, 2000; Manning *et al.*, 1988, 1990, 1995). Several growth factors, growth factor receptors, extracellular proteins, immediate-early genes and cell cycle regulators that may have effects on mammary carcinogenesis have been identified as potential targets of the estrogen-signaling pathway (de Cupis and Favoni, 1997; Katzenellenbogen *et al.*, 1997). However, none of these induced genes can fully explain the mitogenic effects of estrogen or the chemopreventive effects of tamoxifen. In addition, there is an increasing need to identify new estrogen and tamoxifen targets that could be used as biomarkers to monitor the efficacy of cancer treatment and prevention.

In order to determine the global cellular response of breast cancer cells to estrogen and tamoxifen, we have generated SAGE (Serial Analysis of Gene Expression) libraries from an estrogen dependent human breast cancer cell line (ZR75-1) prior to and following estrogen or tamoxifen treatment. SAGE enabled us

*Correspondence: K Polyak, Dana-Farber Cancer Institute, 44 Binney Street D740C, Boston, MA 02115, USA;
E-mail: Kornelia_Polyak@dfci.harvard.edu
Received 15 August 2001; revised 16 October 2001; accepted 29 October 2001

to determine the absolute abundance of thousands of different mRNAs simultaneously in a comprehensive and unbiased way and to detect even slight differences in expression levels between samples (Velculescu *et al.*, 1995). ZR75-1 cells cultured in the absence of endogenous hormones for 7 days were switched to fresh medium in the absence or presence of 10 nM estradiol or 10 μM 4-hydroxy-tamoxifen. Cells were collected after 16 h and response to the hormonal treatment was confirmed by FACS analysis of cell cycle progression and by Northern blot analysis using known estrogen target genes. Estrogen deprived ZR75-1 arrested in G1 and G2 phases of the cell cycle, while addition of estrogen, and to a lesser degree tamoxifen, stimulated rapid S phase entry (data not shown). SAGE libraries from untreated, estrogen or tamoxifen treated cells were generated using a modified micro-SAGE protocol (Porter *et al.*, 2001). From the three SAGE libraries 140 638 tags were obtained, approximately 45 000 from each library (SAGE data will be deposited at http://www.ncbi.nlm.nih.gov/SAGE/). Pair-wise comparison and statistical analysis of these libraries led to the identification of several estrogen and/or tamoxifen induced transcripts. There were 61 tags (33 up-regulated and 22 down-regulated) that showed at least twofold difference ($P<0.001$) between the estrogen treated and control libraries, while 15 tags (nine up-regulated and six down-regulated) showed at least twofold difference ($P<0.001$) between the tamoxifen treated and control libraries. In addition, we found 22 tags that were significantly elevated in the estrogen treated cells when compared to the tamoxifen treated ones, while 24 tags were significantly elevated in the tamoxifen treated library compared to the estrogen treated one. Linking the UniGene database to our SAGE data identified the cDNAs corresponding to the SAGE tags in most of the cases (Table 1). Genes were named according to their abundance in the three SAGE libraries: EIT (Estrogen Induced Tag)-induced by estrogen, TIT (Tamoxifen Induced Tag) -induced by tamoxifen, DET (Differently Expressed Tag)-differently expressed following estrogen or tamoxifen treatment.

Since SAGE tag numbers reflect the absolute abundance of the mRNAs, data obtained from different experiments performed in different laboratories are directly comparable (Velculescu *et al.*, 1995). Therefore, our data and SAGE libraries generated from untreated or estrogen treated MCF-7 cells by others (Charpentier *et al.*, 2000) were analysed using a clustering algorithm to delineate similarities and differences between the effects of estrogen in two different breast cancer cell lines (Figure 1a). Correlating with previous studies the two breast cancer cell lines had distinct genes expression patterns and demonstrated a discrete transcriptional response to estrogen treatment (Perou *et al.*, 2000). Among the known estrogen target genes, cathepsin D was induced by estrogen in both cells lines, whereas pS2 and cyclin D1 were induced only in MCF-7 cells (Figure 1a). All three SAGE libraries derived from the same cell line were more similar to each other, even after estrogen treatment, than to the other cell line. Interestingly, untreated and estrogen treated MCF-7 cells were highly similar to each other, while tamoxifen treated and untreated ZR75-1 cells were somewhat more similar to each other and distinct from estrogen treated cells. These findings indicate that estrogen exerts a differing effect depending on the cellular context, and overall there are relatively few genes significantly affected by estrogen or tamoxifen treatment in these breast cancer cell lines.

To validate the result of the SAGE experiment we have generated probes corresponding to some of the cDNA clones and confirmed their induction by Northern blot analysis (Figure 1b). From the 20 estrogen or tamoxifen induced genes (Table 1) only one (EIT-10=cathepsin D) had been implicated as a target of ER-transcriptional activation, and three had not previously been described at all. In most of the cases the sequences of the genes provided important clues as to their potential functions (Table 1). In particular, several of these genes are predicted to be involved in the regulation of cell proliferation and/or survival. EIT-2 is a protein translocase involved in importing nuclear encoded proteins to the mitochondria (Bauer *et al.*, 1999); EIT-4 is a human homologue of the yeast Dim1p gene essential for mitosis (Berry and Gould, 1997), EIT-6 is an EST homologous to a rat immediate-early gene SM-20 (Wax *et al.*, 1994); TIT-5 is an anti-apoptotic member of the bcl-2 family (Xu and Reed, 1998); and DET-15 and DET-16 are both putative transcription factors with anti-proliferative activity (Ismail *et al.*, 1999; Nakashiro *et al.*, 1998; Shibanuma *et al.*, 1992). Although several ESTs have no homology to known genes, their expression pattern in other SAGE libraries suggests that they also might play a role in cell proliferation and/or estrogen mediated responses (Lal *et al.*, 1999). For example, the TIT-3 and TIT-1 genes appear to be elevated in ER+DCIS (Ductal Carcinoma In Situ) compared to corresponding normal mammary epithelium and expressed at much lower levels in other cell types (Porter *et al.*, 2001). Interestingly, one of the tamoxifen induced genes, SULT1A phenol sulfotransferase, is an enzyme involved in the metabolism of environmental carcinogenes and steroid hormones including estrogen and tamoxifen (Weinshilboum *et al.*, 1997). Recently we and others reported that polymorphism in SULT1A1 influences the age-of onset and the risk of breast cancer, respectively (Seth *et al.*, 2000); Zheng *et al.*, 2001).

Although our SAGE analysis led to the isolation of several novel estrogen and/or tamoxifen regulated genes, it provided no information on which of these genes might be key mediators of the cellular response initiated by estrogen and/or tamoxifen. However, one of the estrogen induced genes, EIT-6, appeared to be particularly interesting due to its relatively high abundance in SAGE libraries prepared from hormone responsive tissues (normal and cancerous mammary,

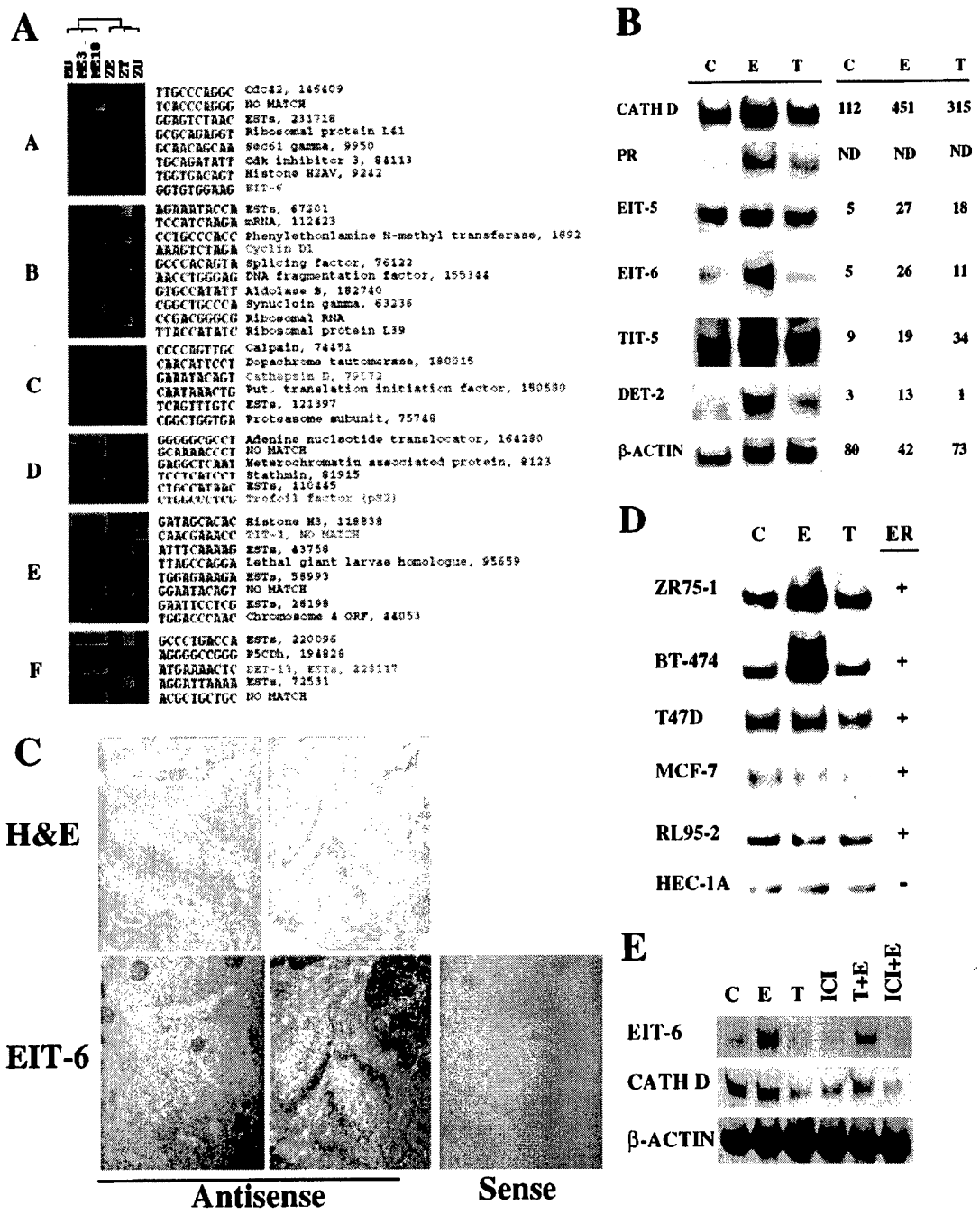


Figure 1 Expression analysis of novel estrogen and tamoxifen induced genes. (**a**) Variation in expression of 2818 genes in six SAGE libraries and dendogram representing similarities in expression pattern among samples. Only parts of certain clusters are included in the figure. These clusters are genes more abundant in ZR75-1 (A) or MCF-7 (B) cells, genes with fairly equal levels in both cell lines and libraries (C), genes induced by estrogen either in ZR75-1 or MCF-7 cells (D) and genes more abundant in tamoxifen and estrogen treated (E) or tamoxifen treated and untreated ZR75-1 (F) cells. Each row represents a gene/SAGE tag, while each column corresponds to a SAGE library. MU, ME3, and ME10 denote MCF-7 cells untreated, estrogen treated for 3, or 10 h, respectively. Similarly ZU, ZE, and ZT stand for untreated, estrogen, or tamoxifen treated ZR75-1 cells. The absolute abundance of the SAGE tag in the library (SAGE tag number) correlates with red color intensity (black = not present – intense red = highly abundant). Genes highlighted in red correspond to known estrogen targets or genes statistically significantly affected by hormonal treatment. (**b**) Correlation of Northern blot (left panel) and SAGE (right panel; results for the indicated genes. Numbers represent SAGE tag numbers. RNA and SAGE libraries were prepared from untreated (C), estrogen (E) and tamoxifen (T) treated cells. CATH D and PR denotes cathepsin D and progesterone receptor, respectively. (**c**) *In vivo* expression of EIT-6 in normal breast tissue. Histological sections of normal human breast tissue were stained with hematoxylin-eosin (H&E). Adjacent slides were hybridized with P^{33} labeled EIT-6 anti-sense or sense probes and visualized using dark-field microscopy. Intense hybridization signal is detected in mammary epithelial cells using the antisense, but not the sense probe. (**d**) Analysis of EIT-6 mRNA levels in various

ovarian and prostate epithelium) (Lal *et al.*, 1999; Lash *et al.*, 2000), and we therefore characterized it in further detail. A full-length (2071 bp) human EIT-6 cDNA was obtained by analysing and sequencing ESTs clones. The human EIT-6 cDNA contains an open reading frame (ORF) of 1221 bp encoding a predicted protein of 407 amino acids (~43 000 Daltons), which was confirmed by *in vitro* transcription/translation experiments (data not shown). By FISH analysis we localized EIT-6 to chromosome 19q13.1, a region not previously implicated in breast cancer (data not shown). We also analysed the *in vivo* abundance of the EIT-6 mRNA and determined its expression at the cellular level by mRNA *in situ* hybridization of normal human breast tissue (Figure 1c). EIT-6 hybridization signal showed fairly even intensity throughout the mammary epithelium, while no significant signal was detected in stromal cells.

To determine how generally EIT-6 is regulated by estrogen, we performed Northern blot analysis of various ER+breast and endometrial cancer cell lines (Figure 1d). In addition to being induced in ZR75-1 cells (~5-fold induction), the cell line used for the generation of the SAGE libraries, EIT-6 is induced in BT-474 cells (~10-fold induction), but not in other estrogen responsive cell lines analysed. This finding is not unexpected, since many estrogen targets are induced in a cell type specific manner.

To determine if the induction of EIT-6 by estrogen is ER mediated, we analysed the effect of estrogen antagonists (4-hydroxy-tamoxifen and ICI 182 780) on EIT-6 mRNA levels (Figure 1e). Consistent with our SAGE data, EIT-6 was induced only by estrogen, and this induction was completely or partially abolished by the addition of ICI 182 780 or tamoxifen, respectively. Thus, the increase in EIT-6 mRNA levels following estrogen treatment is both estrogen and ER dependent. There are multiple ways estrogen may be regulating the expression levels of EIT-6: (1) direct transcriptional regulation; (2) influencing mRNA stability; or (3) indirectly through other transcription factors/signaling pathways. Northern blot analysis of EIT-6 mRNA levels at different time points following estrogen treatment indicated that EIT-6 is induced by estrogen at about the same time as cathepsin D, a known direct estrogen target, (Figure 2a). This result suggests that EIT-6, similar to cathepsin D, may also be a direct transcriptional target of ER. Addition of transcription inhibitors completely abolished EIT-6 induction by estrogen indicating that increased EIT-6 mRNA levels are likely due to increased transcription (data not shown). To further investigate if EIT-6 is a

Table 1 Estrogen and tamoxifen regulated genes

Name	*SAGE tag*	*C*	*E*	*T*	*GenBank#*	*Function*
EIT-1	GCGGTGACAG	1	14	8	NA	No reliable database match
EIT-2	TACGAAGTTC	1	14	9	AF077039	TIM17B
EIT-3	AATGAGTTTG	2	19	13	AF201940	DC6 mRNA
EIT-4	GTCTTAACTC	2	18	9	BC001046	Yeast Dim1 homologue
EIT-5	GTGGCATCAC	5	27	18	AB043104	No10p snRNP
EIT-6	GGTGTGGAAG	5	26	11	AY040565	SM-20 homologue
EIT-10	GAAATACAGT	112	451	315	X05344	Cathepsin D ECM protease
TIT-1	CAACGAAACC	0	4	14	NA	No reliable database match
TIT-2	GCGTGCTCTC	0	3	11	NA	No reliable database match
TIT-3	GGGGGCCCCG	8	28	32	AF004876	Yeast Yif1p homologue
TIT-4	GGGGCCCCCT	7	27	28	Z96932	Sjogren's syndrome antigen
TIT-5	CCACCCCGAA	9	19	34	BC000916	BI-I bax antagonist
DET-1	TCTCTGCAAA	4	15	1	BC000890	Hypothetical protein FLJ20640
DET-2	TGGATCCTCG	3	13	1	BC001239	Hypothetical protein FLJ10479
DET-13	ATGAAAACTC	7	1	11	AA524901	ESTs no homology
DET-14	AGCCACCGTG	4	1	12	NA	No reliable database match
DET-15	CCCCCGCGGA	11	1	12	AF130366	USF2 transcription factor
DET-16	TTTGCGGTCC	7	0	11	BC002972	TSC-22-like protein
DET-17	GCTGGGGACT	0	1	10	NM_001055	SULT 1A sulfotransferase

breast (ZR75-1, BT-474, T47D, MCF-7) and endometrial (RL95-2, HEC-1A) cancer cell lines following estrogen (E) or tamoxifen (T) treatment. Estrogen receptor (ER) status of the cell lines are indicated by '+' and '−' signs. (e) EIT-6 expression in cells treated with no drug (C), estrogen (E), tamoxifen (T), ICI 128 780 (ICI) or with combinations of estrogen and tamoxifen (E+T) and estrogen and ICI (E+ICI). Breast cancer cell lines were obtained from American Type Culture Collection and maintained according to the supplier. To assay estrogen responsiveness cells were cultured in phenol red free medium (Life Technologies) supplemented with 5% charcoal treated fetal bovine serum (Hyclone) after which cells were switched to fresh medium or fresh medium containing 10 nM estradiol or 10 μM 4-hydroxy-tamoxifen. SAGE libraries were generated and analysed as previously described (Porter *et al.*, 2001). Hierarchical clustering was applied to data using the Cluster program developed by Eisen *et al.* (1998). Data was log-transformed and filtered for at least one observations abs Val 5 and Maxval-Minval>5. Using these settings 2818 genes (out of 16 808 total) were included in the analysis. Results were displayed with the TreeView program (Eisen *et al.*, 1998). mRNA *in situ* hybridizations using a P^{33} labeled sense or anti-sense EIT-6 ribo-probes were performed as described (Rosen *et al.*, 1999). The effect of estrogen antagonists was determined by pre-incubating the cells with 1 μM ICI 128 780 or 10 μM 4-hydroxy-tamoxifen for 6 h followed by estrogen treatment for an additional 24 h. RNA isolation, RT–PCR and Northern blot analyses were performed as described (Polyak *et al.*, 1997)

direct or indirect target of ER we analysed its mRNA levels following estrogen treatment in the presence of a protein synthesis inhibitor (cycloheximide) in BT-474 cells (Figure 2b). Cycloheximide and estrogen both increased EIT-6 mRNA levels to a certain degree, but estrogen treatment in the presence of cycloheximide led to a much stronger induction. Based on these results, the induction of EIT-6 following estrogen treatment does not appear to require new protein synthesis. Therefore, EIT-6 is likely to be a direct transcriptional target of ER, but the possibility that other proteins are also involved in the transcriptional activation of EIT-6 cannot be excluded.

To further characterize the mechanism by which ER induces EIT-6, first we determined if a ~5.5 kb fragment of the proximal EIT-6 promoter confers estrogen responsiveness to a luciferase reporter gene. Measurement of luciferase activity in 911 cells following transient transfection of this construct with co-transfected ERα demonstrated a modest, but reproducible induction following estrogen treatment (Figure 2c). Random 2–3 kb fragments derived from the EIT-6 genomic region placed up-stream of a luciferase gene conferred no response to estrogen treatment (data not shown). Next, we analysed the sequence of this promoter region and identified several potential estrogen responsive elements (ERE) closely resembling the consensus ERE sequence (Figure 2d). To determine whether these putative EREs can confer estrogen responsiveness, we generated various constructs with concatemers of each of these elements or their combination placed up-stream of a luciferase gene (Figure 2d). Measurement of luciferase activity following transient transfection of these constructs with co-

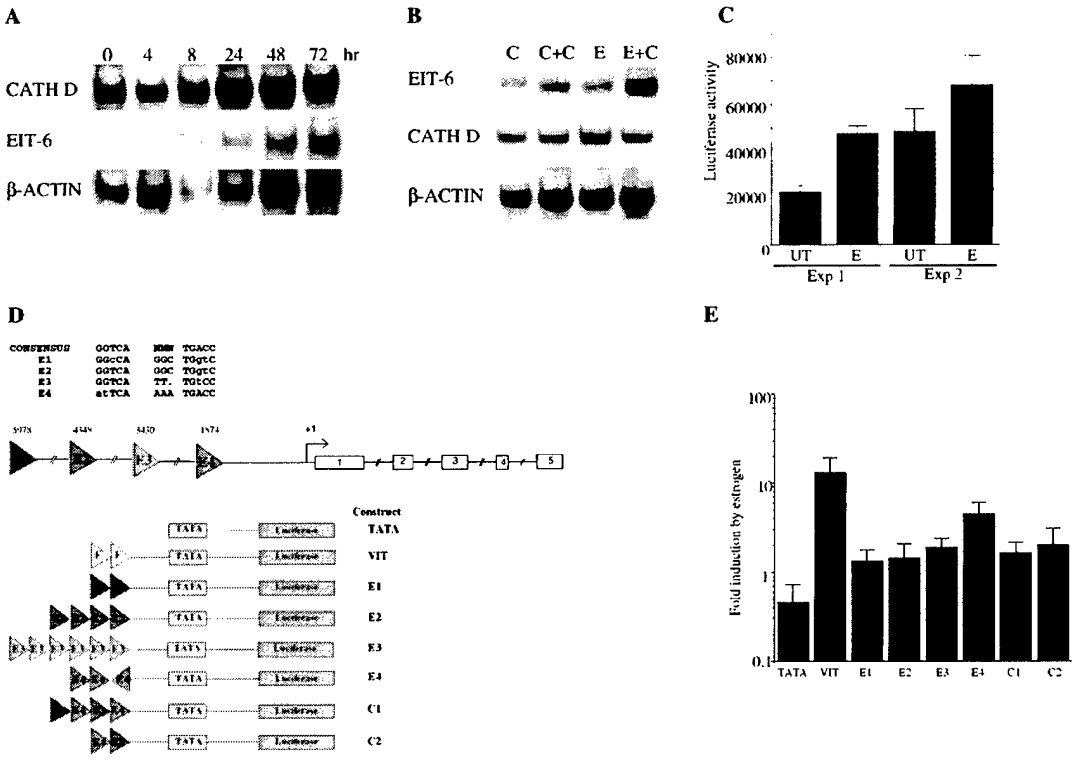


Figure 2 EIT-6 is a direct estrogen receptor target. (**a**) Northern blot analysis of time-course of EIT-6 and Cathepsin D (CATH D) induction following estrogen treatment for the indicated time. (**b**) Analysis of the effect of cycloheximide (CHX) on EIT-6 mRNA levels. Cells were treated with ethanol alone (C) or together with CHX (C+C), and estrogen alone (E) or together with CHX (E+C). (**c**) Representative experiments demonstrating increased luciferase activity following estrogen treatment in cells transiently transfected with an EIT-6 5.5 kb promoter luciferase construct. UT and E denote untreated and estrogen treated cells, respectively. (**d**) EIT-6 genomic structure and reporter constructs. (**e**), Results of luciferase assays following transient transfections of HepG2 cells with the indicated constructs (x-axis). Fold induction by estrogen is indicated on y-axis. Numbers are average of three independent experiments performed in quadruplicate, luciferase activity was normalized for transfection efficiency by using the ratio of luciferase to β-galactosidase activity. A construct containing two copies from the vitillogenin. A promoter was used as positive control (VIT) (McMahon *et al*., 1999). Estrogen treatment, RNA isolation and Northern blot analysis was performed as described above. The effect of cycloheximide (CHX) was assayed by treating the cells with 10 μg/ml CHX for 16 h in the presence or absence of estrogen. EIT-6 promoter luciferase reporter constructs were generated by subcloning a ~5.5 kb BAC (bacterial artificial chromosome) derived fragment of the EIT-6 promoter or concatemers of PCR-derived fragments of the human EIT-6 promoter containing the putative EREs into pBR-pl-luc or pBR-pl-TATA-luc, respectively (Polyak *et al*., 1997). Cells were transfected using FuGene6 (Roche), treated with estrogen the day after transfection and the following day luciferase and β-galactosidase activities were determined using a luciferase assay system (Promega) and the Aurora GAL-XE reporter gene assay (ICN), respectively

transfection of ERα in HepG2 cells revealed that most of them demonstrated some, although relatively weak, estrogen responsiveness (Figure 2e). The most significant (4–5 fold) induction was observed using the E4 construct containing three copies of the E4 ERE. This ERE is closest to the transcription start site and contains a nearly perfect ERE with only two mismatches compared to the consensus sequence. In the same experiments two copies of the consensus ERE derived from the vitillogenin promoter (VIT) led to a 10–11-fold induction in luciferase activity following estrogen treatment (Figure 2e). Therefore, a 4–5-fold induction observed with the E4 ERE is significant and indicates that this fragment could be a functional ERE. Although these experiments do not prove that any of these putative EREs are necessary for the induction of EIT-6 by estrogen, they do show that these elements can confer estrogen responsiveness. Therefore, direct binding of ER to these putative EREs may be responsible for the transcriptional induction of EIT-6, but other mechanisms cannot be excluded.

EIT-6 is a novel human gene and although it is homologous to the SM-20 rat immediate-early gene induced by growth agonists (Wax *et al.*, 1994), it is not

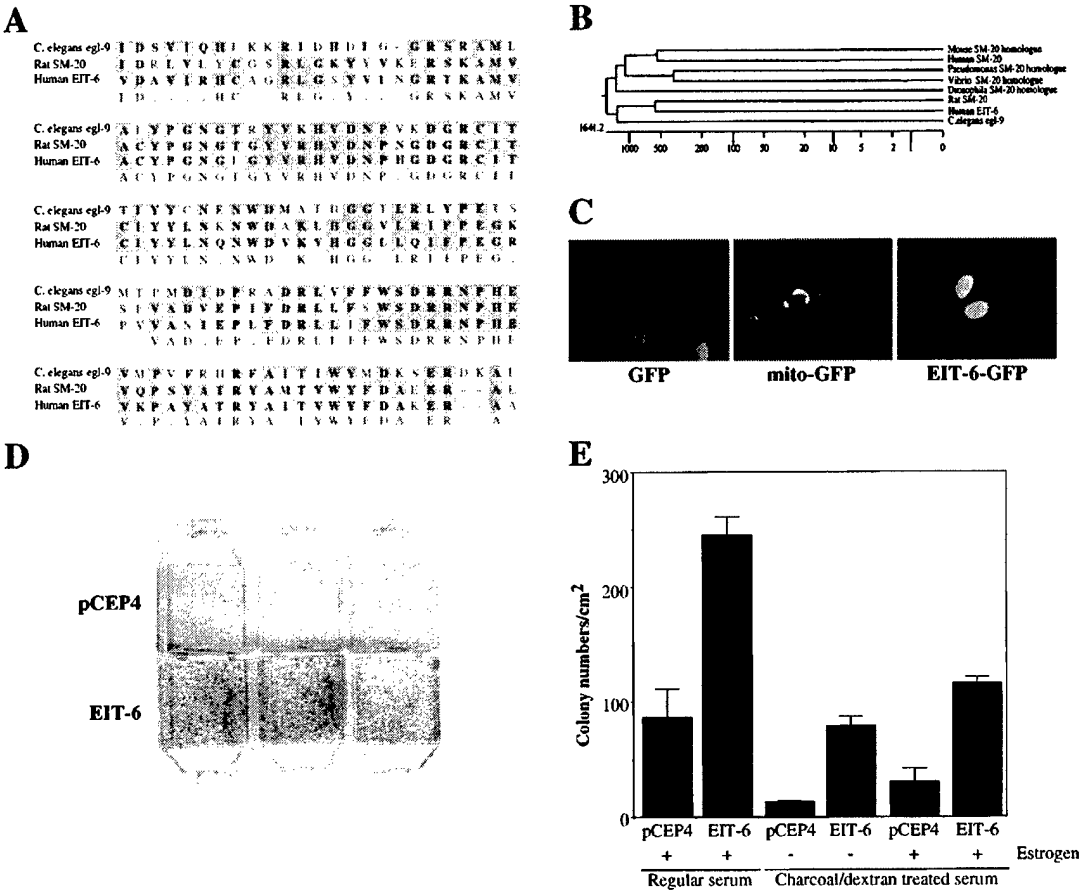


Figure 3 EIT-6: homologues and putative function. (**a**) Amino acid alignment of C-terminal parts of human EIT-6 (amino acids 260–383), rat SM20 (amino acids 214–336) and *C. elegans* egl-9 (amino acids 451–574) proteins using MacVector and ClustalW alignment. (**b**) Phylogenetic comparison of EIT-6 homologues. Comparisons were made using DNAStar and the Jotun Hein algorithm. (**c**) Subcellular localization of EIT-6. Cells transfected with constructs encoding control GFP and mito-GFP or EIT-6-GFP proteins were visualized with fluorescence microscopy. Magnification objective 20×. (**d**) Representative flasks from colony assay experiments. (**e**) Quantification of colony assay experiments. Presence (+) or absence (−) of estrogen and plasmids transfected (pCEP4 or pCEP4-EIT6) are indicated on the x axis, while colony numbers (per cm^2) are plotted on the y axis. Results are the average of three independent experiments. For intracellular localization studies EIT-6-GFP fusion fragment was PCR amplified as described (Flatt *et al.*, 2000) and subcloned into pShuttle-CMV construct (He *et al.*, 1998). Cells were transfected with pEGFP-N, pEGFP-mito (Clonetech) or pShuttle-EIT-6-GFP plasmids and analysed by fluorescence microscopy. To assay colony growth expression construct was generated by subcloning a PCR derived EIT-6 cDNA with a C-terminal double hemagglutinin (HA2) tag into pCEP4 (Invitrogen). T47D cells were transfected with FuGene6 (Roche) and selected in hygromycin containing medium for 2 weeks after which colonies were visualized by crystal violet staining. The number of colonies/cm^2 was determined based on spot-densitometry assisted counting using a MultiImage Lightbox (Alfa Innotech). Two independent areas/flask and two flasks/experiments were analysed. In some experiments the cells were grown in regular 10% fetal bovine serum (FBS) containing RPMI medium, while in other experiments cells were cultured in phenol red free RPMI medium (Life Technologies) supplemented with 5% charcoal treated FBS (Hyclone) in the presence or absence or 10 nM estradiol

the human orthologue of this rat gene (Figure 3a). However, similar to SM-20, EIT-6 is also induced by various growth agonists (EGF, isoproterenol, and PMA) in ZR75-1 cells (data not shown). The rat SM-20 cDNA was also identified as a gene induced by wild-type p53 and as a gene induced in sympathetic neurons during NGF (Nerve Growth Factor) withdrawal initiated apoptosis (Lipscomb *et al.*, 1999; Madden *et al.*, 1996). Subsequent experiments in muscle cells determined that SM-20 may play a role in the regulation of myoblast proliferation and differentiation (Moschella *et al.*, 1999). Interestingly, a recent study isolated another SM-20/EIT-6 homologue as a gene significantly overexpressed in over 50% of endometrial carcinomas, the development of which is thought to depend on estrogenic hormones (Foca *et al.*, 2000). Based on these results we can conclude that EIT-6 is a member of a multi gene family of growth regulatory genes that includes SM-20, EIT-6, and additional SM-20 homologues. EIT-6 appears to encode a protein with evolutionarily conserved function: there is a *C. elegans* EIT-6 homologue (Figure 3a,b) that was identified as an egg laying defective mutant (egl-9) (Trent *et al.*, 1983), and several additional SM-20/EIT-6 homologues were identified from various other species including several types of bacteria (Figure 3b).

Surprisingly immunohistochemical analysis of the rat SM-20 gene demonstrated cytoplasmic staining, while the human SM-20 orthologue, another related human gene (SCAND2), and EIT-6 all contain putative nuclear localization signals (Dupuy *et al.*, 2000; Wax *et al.*, 1994). However, most of the homology between EIT-6 and SM-20 resides in the C-terminal region, while the N-terminal domain (containing the nuclear localization sequence) are divergent. Therefore, to determine the sub-cellular localization of EIT-6, we generated a construct expressing an EIT-6-GFP (Green Fluorescence Protein) fusion protein. Fluorescence microscopic analysis of cells transiently transfected with control GFP and mitochondrial-GFP (mito-GFP) encoding constructs revealed mostly cytoplasmic and mitochondrial localization, respectively (Figure 3c). In contrast, the EIT-6-GFP protein was detected only in the nucleus. Similar results were obtained by Western blot analysis of fractionated cell extracts prepared from cells expressing a hemaglutinine epitope tagged EIT-6 protein (data not shown). We were unable to determine the localization of the endogenous EIT-6 protein, since the polyclonal antibodies we generated were not suitable for immunohistochemical analysis. However, based on these data EIT-6 is likely to be located and function in the nucleus.

Based on its homology to SM-20 we hypothesized that EIT-6 expression influences cell proliferation and/or survival. In order to test this, we transfected T47D ER+breast cancer cells with control (pCEP4) or pCEP4-EIT6 (encoding a C-terminal hemaglutinin epitope tagged EIT-6 protein) constructs. Stable transfectants were selected by culturing the cells in the presence of hygromycin for 2 weeks after which colonies were visualized by crystal violet staining. The expression of EIT-6 was confirmed by immunoblot analysis of cell extracts prepared from pools of stable clones using anti-HA antibody (data not shown). In three independent experiments performed in duplicate flasks, expression of EIT-6 led to a significant (3–4-fold) increase in colony numbers (representative examples Figure 3d, summary of colony counts in Figure 3e). Similar results were obtained in MDA-MB-435S ER negative breast cancer cells (data not shown). These data indicate that EIT-6 overexpression enhances colony growth in human breast cancer cells.

To test if expression of EIT-6 can confer estrogen independent growth, we transfected the ER+and estrogen dependent T47D breast cancer cells with control pCEP4 or pCEP4-EIT6 constructs. Stable transfectants were selected by culturing the cells in the absence of hormones in phenol red free medium supplemented with 5% charcoal/dextran treated fetal bovine serum and hygromycin. Half of the flasks were cultured in the presence of 10 nM estrogen. Colonies were visualized by crystal violet staining after 2 weeks of selection. Very few colonies were observed in the absence of estrogen in control pCEP4 transfected cells confirming the requirement of estrogen for T47D cell growth (Figure 3e). In contrast, a significant number of colonies was observed in EIT-6 transfected cells in the absence of estrogen indicating that EIT-6 expression relieves estrogen dependency (Figure 3e). Addition of estrogen increased colony numbers in both control pCEP4 and EIT-6 transfected cells, therefore, EIT-6 expression may not be sufficient to completely alleviate estrogen dependence (Figure 3e). Due to the lack of suitable antibodies we were unable to determine the relative levels of ectopic and endogenous EIT-6 proteins in these clones. Therefore, the possibility that increased colony numbers were due to exogenous EIT-6 protein levels significantly exceeding endogenous estrogen induced EIT-6 protein levels cannot be excluded.

In summary, based on SAGE analysis of estrogen and tamoxifen treated breast cancer cells, we identified several novel putative targets of the estrogen signaling pathway. One of these genes, EIT-6, may be involved in transmitting growth promoting signals initiated by estrogen in human breast cancer cells. EIT-6 has no known functional domains and EIT-6 homologues are not well characterized, therefore, we can only speculate on how EIT-6 may regulate estrogen dependent and independent cell growth. Since EIT-6 is a nuclear protein one attractive hypothesis is that EIT-6 may influence the transcription of genes involved in the regulation of cell proliferation or survival. However, further studies are required to determine if EIT-6 or any of the other genes are critical downstream components of the estrogen signaling pathway.

Note added in proof

During the review of this manuscript two independent studies (Epstein *et al.*, *Cell*, **107**, 43–54 and Bruick *et al.*, *Science*, **294**, 1337–1340) identified EIT-6 as a dioxygenase that may regulate HIF (hypoxia inducible factor) by hydroxylation.

Acknowledgments
We thank Carmen Tam and Massimo Loda for mRNA *in situ* hybridization. Jim DiRenzo and Yongfeng Shang for helpful advice with luciferase assays. Danielle Garneau and Ed Fox for sequencing the SAGE libraries, Bill Sellers, Myles Brown, and Alexander Miron for critical review of the manuscript. This work was, in part, supported by the Sydney Kimmel Cancer Foundation (K Polyak), a US department of Defense postdoctoral fellowship (DAMD17-01-1-0221) to P Seth and the Dana-Farber Cancer Institute.

Accession number
The GenBank accession number for the human EIT-6 cDNA sequence reported in this paper is AY040565.

References

Bauer MF, Gempel K, Reichert AS, Rappold GA, Lichtner P, Gerbitz KD, Neupert W, Brunner M and Hofmann S. (1999). *J. Mol. Biol.*, **289,** 69–82.
Berry LD and Gould KL. (1997). *J. Cell. Biol.*, **137,** 1337–1354.
Charpentier A, Bednarek A, Daniel R, Hawkins K, Laflin K, Gaddis S, MacLeod M and Aldaz C. (2000). *Cancer Res.*, **60,** 5977–5983.
de Cupis A and Favoni RE. (1997). *Trends Pharmacol. Sci.*, **18,** 245–251.
Dupuy D, Aubert I, Duperat VG, Petit J, Taine L, Stef M, Bloch B and Arveiler B. (2000). *Genomics,* **69,** 348–354.
Eisen MB, Spellman PT, Brown PO and Botstein D. (1998). *Proc. Natl. Acad. Sci. USA,* **95,** 14863–14868.
Fisher B, Costantino JP, Wickerham DL, Redmond CK, Kavanah M, Cronin WM, Vogel V, Robidoux A, Dimitrov N, Atkins J, Daly M, Wieand S, Tan-Chiu E, Ford L and Wolmark N. (1998). *J. Natl. Cancer Inst.*, **90,** 1371–1388.
Fisher B, Dignam J, Wolmark N, Wickerham DL, Fisher ER, Mamounas E, Smith R, Begovic M, Dimitrov NV, Margolese RG, Kardinal CG, Kavanah MT, Fehrenbacher L and Oishi RH. (1999). *Lancet,* **353,** 1993–2000.
Flatt PM, Polyak K, Tang LJ, Scatena CD, Westfall MD, Rubinstein LA, Yu J, Kinzler KW, Vogelstein B, Hill DE and Pietenpol JA. (2000). *Cancer Lett.*, **156,** 63–72.
Foca C, Rice GE, Quinn MA and Moses EK. (2000). *Mol. Hum. Reprod.*, **6,** 712–718.
Freedman LP. (1999). *Trends Endocrinol. Metab.*, **10,** 403–407.
Greenlee RT, Murray T, Bolden S and Wingo PA. (2000). *CA Cancer J. Clin.*, **50,** 7–33.
He TC, Zhou S, da Costa LT, Yu J, Kinzler KW and Vogelstein B. (1998). *Proc. Natl. Acad. Sci. USA,* **95,** 2509–2514.
Inadera H, Hashimoto S, Dong HY, Suzuki T, Nagai S, Yamashita T, Toyoda N and Matsushima K. (2000). *Biochem. Biophys. Res. Commun.*, **275,** 108–114.
Ismail PM, Lu T and Sawadogo M. (1999). *Oncogene,* **18,** 5582–5591.
Katzenellenbogen BS. (1996). *Biol. Reprod.*, **54,** 287–293.
Katzenellenbogen BS, Montano MM, Ekena K, Herman ME and McInerney EM. (1997). *Breast Cancer Res. Treat.*, **44,** 23–38.
Lal A, Lash AE, Altschul SF, Velculescu V, Zhang L, McLendon RE, Marra MA, Prange C, Morin PJ, Polyak K, Papadopoulos N, Vogelstein B, Kinzler KW, Strausberg RL and Riggins GJ. (1999). *Cancer Res.*, **59,** 5403–5407.
Lash AE, Tolstoshev CM, Wagner L, Schuler GD, Strausberg RL, Riggins GJ and Altschul SF. (2000). *Genome Res.*, **10,** 1051–1060.
Lipscomb EA, Sarmiere PD, Crowder RJ and Freeman RS. (1999). *J. Neurochem.*, **73,** 429–432.
Madden SL, Galella EA, Riley D, Bertelsen AH and Beaudry GA. (1996). *Cancer Res.*, **56,** 5384–5390.
Mangelsdorf DJ, Thummel C, Beato M, Herrlich P, Schutz G, Umesono K, Blumberg B, Kastner P, Mark M, Chambon P and Evans R. (1995). *Cell,* **83,** 835–839.
Manning DL, Archibald LH and Ow KT. (1990). *Cancer Res.*, **50,** 4098–4104.
Manning DL, Daly RJ, Lord PG, Kelly KF and Green CD. (1988). *Mol. Cell. Endocrinol.*, **59,** 205–212.
Manning DL, McClelland RA, Knowlden JM, Bryant S, Gee JM, Green CD, Robertson JF, Blamey RW, Sutherland RL, Ormandy CJ and Nicholson RI. (1995). *Acta Oncol.*, **34,** 641–646.
McMahon C, Suthiphongchai T, DiRenzo J and Ewen ME. (1999). *Proc. Natl. Acad. Sci. USA,* **96,** 5382–5387.
Moschella MC, Menzies K, Tsao L, Lieb MA, Kohtz JD, Kohtz DS and Taubman MB. (1999). *Gene Expr.*, **8,** 59–66.
Nakashiro K, Kawamata H, Hino S, Uchida D, Miwa Y, Hamano H, Omotehara F, Yoshida H and Sato M. (1998). *Cancer Res.*, **58,** 549–555.
Nandi S, Guzman RC and Yang J. (1995). *Proc. Natl. Acad. Sci. USA,* **92,** 3650–3657.
Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA, Fluge O, Pergamenschikov A, Williams C, Zhu SX, Lonning PE, Borresen-Dale AL, Brown PO and Botstein D. (2000). *Nature,* **406,** 747–752.
Polyak K, Xia Y, Zweier JL, Kinzler KW and Vogelstein B. (1997). *Nature,* **389,** 300–305.
Porter D, Krop I, Nasser S, Sgroi D, Kaelin C, Marks J, Riggins G and Polyak K. (2001). *Cancer Res.*, **61,** 5697–5702.
Rosen ED, Sarraf P, Troy AE, Bradwin G, Moore K, Milstone DS, Spiegelman BM and Mortensen RM. (1999). *Mol. Cell.*, **4,** 611–617.
Seth P, Lunetta KL, Bell DW, Gray H, Nasser SM, Rhei E, Kaelin CM, Iglehart DJ, Marks JR, Garber JE, Haber DA and Polyak K. (2000). *Cancer Res.*, **60,** 6859–6863.
Shibanuma M, Kuroki T and Nose K. (1992). *J. Biol. Chem.*, **267,** 10219–10224.
Trent C, Tsung N and Horvitz R. (1983). *Genetics,* **104,** 619–647.
Velculescu VE, Zhang L, Vogelstein B and Kinzler KW. (1995). *Science,* **270,** 484–487.
Wax SD, Rosenfield CL and Taubman MB. (1994). *J. Biol. Chem.*, **269,** 13041–13047.
Weinshilboum RM, Otterness DM, Aksoy IA, Wood TC, Her C and Raftogianis RB. (1997). *FASEB J.*, **11,** 3–14.
Xu Q and Reed JC. (1998). *Mol. Cell.*, **1,** 337–346.
Zheng W, Xie D, Cerhan JR, Sellers TA, Wen W and Folsom AR. (2001). *Cancer Epidemiol. Biomarkers Prev.*, **10,** 89–94.